1982

Behavioral Consistency And Individual Differences In Predictive Structure

Sampo Vilho Paunonen

CANADIAN THESES ON MICROFICHE

I.S.B.N.

THESES CANADIENNES SUR MICROFICHE

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ MICROFILMÉE TELLE QUE NOUS L'AVONS REÇUE

BEHAVIOURAL CONSISTENCY AND INDIVIDUAL DIFFERENCES IN PREDICTIVE STRUCTURE

by

Sampo Vilho Paunonen

Department of Psychology

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
July, 1982

ABSTRACT

Psychological research has frequently been directed at evaluating the predictive utility of personality assessments with respect to various behavioural criteria. Bem and Allen (1974) have advocated an approach to maximizing the convergences of various modes of personality assessment which is based on the identification of behaviourally consistent and predictable individuals along specific trait dimensions. The measured behavioural consistency for individuals was used as an independent moderator variable, influencing the correlations between predictor and criterion measures of personality. Individuals who identified themselves as being consistent for a particular behavioural domain were shown by Bem and Allen to be more predictable than less consistent subjects with regard to peer ratings and other trait measures.

Subsequent research by Kenrick and Stringfield (1980) and others (e.g., Kenrick & Braver, 1982) has provided some support for the idiographic formulations of Bem and Allen. Procedural differences among empirical studies, however, have produced generally unrecognized problems of interpretation related to the outcomes of hypothesis tests. These problems are described and evaluated in the present investigation.

The results of a peer rating study of personality, and an assimilation of data from other sources, suggest that (a) behavioural consistency is related to trait level or need strength such that the greatest degrees of consistency occur at the extremes of bipolar dimensions of behaviour, and that (b) the failure to consider this relationship in classifying subjects as predictable (consistent) and unpredictable (variable) can lead to the discovery of moderator effects that are spurious. The consistent subgroups being predominately the most extreme on personality dimensions will have inflated trait score variances that increase the likelihood of finding significant correlational indices of behavioural predictability. Accounting for the relationship between consistency scores and personality scores in this study resulted in little evidence for the pervasive belief that individual differences in trait consistency effectively moderate the validity of behavioural assessments.

Additional analyses presented and discussed are concerned with other issues relevant to the validity of personality assessments. These issues include rater confidence as a moderator of predictive validity, reliability and validity as a function of the aggregation of replicated measurements, and the stability of the measured organization of personality traits.

ACKNOWLEDGEMENTS

I would like to express my sincerest appreciation to Douglas N. Jackson who, as my advisor, has supported me in many ways throughout the completion of this project. I would also like to thank the members of the Advisory and Examining Committees for their contributions to this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Human behaviours are remarkably variegated and their determinants frequently complex. Accordingly, the prediction of distinct individual responses to circumscribed stimuli has met with somewhat less than a desirable level of success. In a highly influential and widely cited review of research in the area, Mischel (1968) concluded that empirical studies designed to evaluate the predictive utility of various assessments of personality have typically yielded inadequate results. Validity coefficients, or the intercorrelations among predictors of the same behaviour, were seldom found to exceed values of approximately .30. Reasons for the plethora of discouraging results have not been wanting (e.g., Endler & Magnusson, 1976; Fiske, 1978; Jackson, 1982; Mischel, 1968; Schneider, 1973; Wiggins, 1973), many of which make reference to errors of measurement, perceptual distortions, misrepresentations in memory, suboptimal prediction strategies, and inattentiveness to person by situation interactions.

An immediate and salutary product of the ensuing controversy has been an impassioned search for optimal conditions fostering maximal validity in personality measurement. One such approach is predicated on the notion that individual rather than general laws are thought to

govern human behaviour; an idea reflecting the idiographic-nomothetic distinction originally highlighted by Allport in 1937. Proponents of an idiographic system of psychology maintain that, if the behaviour of organisms is not determined by global laws, it makes little sense for the scientist to apply an exclusive prediction strategy universally. Rather, the unique determinants of each individual's behaviour require identification before the blueprinting of contingent prediction procedures.

The purpose of the present paper is to evaluate, both logically and empirically, an "idiographic approach" to maximizing behavioural predictability originally proposed by Bem and Allen (1974) and espoused most recently by Bem and Funder (1978), Lamiell (1981), Kenrick and Braver (1982), Kenrick and Stringfield (1980), Schneiderman (1980), and Underwood and Moore (1981). Advocates of this position maintain that there are stable individual differences in the cross-situational regularity of behaviours related to particular construct domains or traits, and the more consistent a person with respect to some class of behaviours, the more predictable is the occurrence of those behaviours. This thesis has been extended to imply that trait related behaviours for highly inconsistent individuals are not predictable. Such traits are believed to be not "relevant" to those individuals in that the constructs are not useful for "characterizing" them (Bem & Allen, 1974, p. 509). This differential

consideration does not distinguish nomothetic conceptualizations where all dimensions of behaviour are conceived of as universally applicable. It has been argued that idiographic measurement strategies will allow for more accurate prediction equations than those based on nomothetic principles. These idiographic personality assessment procedures are designed to enable the identification of consistent and therefore predictable individuals on any particular dimension of behaviour.

The present investigation has several aims. In the remainder of this chapter the origins and implications of the issue of behavioural consistency as an influence on behavioural predictability are examined in detail and critically evaluated. An empirical study, designed to extend the findings of previous investigations while addressing some of their problems, is presented in the following chapters. Discussion is focussed on the relevance of the results of the study to current thinking on the behavioural consistency variable as it pertains to individual differences in predictive structure.

Moderator Variables

The main thrust of the study by Bem and Allen (1974) is that behavioural consistency can be used as a moderator variable to improve the predictability of personality assessments. A moderator variable modifies, or makes conditional, the relationship between two other variables. That is, the association between predictor and criterion

varies as a function of the moderator variable (Saunders, 1956; Wiggins, 1973, p. 53). For example, in one of the earliest formal attempts at enhancing the predictive utility of self report test scores, Frederiksen and Melville (1954) evaluated the hypothesis that a measure of "compulsiveness" would moderate the relationship between an engineering interest measure and grade point average in a sample of freshman engineering students. They argued that the relationship between students' interest in a programme of study and academic achievement would be stronger for students low in compulsiveness whose performance is not impeded by obsessive tendencies. The correlation between interest scores and grade point average was .10 in the full sample of 154 engineering students. If, however, the sample was split into high and low halves based on an index of compulsiveness, the corresponding correlations for the two groups differed substantially; -.01 for the compulsive individuals and .25 for the non-compulsive subgroup. The variable of compulsiveness, in the Frederiksen and Melville study, moderated the relationship between interest scores and academic performance (cf. Kellogg, 1968; Stricker, 1966).

More recently, Scheier, Buss, and Buss (1978) have reported that individuals high in "private" self-consciousness (those who are attentive to personal aspects of the self such as motives and feelings) show a substantially greater correspondence between self reported

aggression and experimentally induced aggressive behaviours compared to those low on the dimension. As hypothesized, "public" self-consciousness (attending to social aspects of the self such as appearance) did not moderate the association between self-reported and objectively measured behaviour.

Such studies are rare examples of successful applications of theoretically derived moderators to improve the predictive utility of behavioural equations. Wiggins (1973, ch. 2) has presented an excellent account of various enterprising projects that have attempted to isolate these seemingly elusive variables.

The Bem and Allen Study

Bem and Allen (1974) have demonstrated an increasing convergence between a variety of modes of behavioural assessment when the moderating effect of self reported behavioural consistency was taken into consideration. Focussing on the personality dimensions of friendliness and conscientiousness, they administered an 86-item multiscale questionnaire to several subjects, containing 24 items related to the first dimension and 23 items keyed for the second. Employed were behaviour-situation exemplars such as "When in a store, how likely are you to strike up a conversation with a sales clerk?" (p. 512). Ratings of peers and parents on the same items were later collected, in addition to more objective behavioural scores such as the frequency and duration of group discussion

(friendliness) and promptness in returning course assignments (conscientiousness). Also collected was each subject's global estimate of behavioural consistency. Individual variability for a trait was assessed by one item (e.g., "How much do you vary from one situation to another in how friendly and outgoing you are?", p. 512). All responses were recorded on 7-point scales.

The ratings and objective measures of behaviours related to friendliness showed substantial positive intercorrelations for those individuals claiming to be highly consistent in their degree of friendliness. This observation was in contrast to the generally lower correlations based on the more variable subjects. Similar results were obtained using the dimension of conscientiousness, although a questionable statistic based on item response variances, described below, was used to separate subjects into high and low variability subgroups. Self reported consistency failed to discriminate predictable from unpredictable individuals on this dimension. These findings have prompted Bem and Allen (1974) to view behavioural consistency as a moderator variable, influencing the correspondence between self ratings of personality and other assessments of the same constructs.

The high and low variability subjects in Bem and Allen's study were matched with respect to their actual position on each of the trait dimensions, as measured by

the 86 item multiscale personality questionnaire. In this way the researchers purportedly experimentally partialled out any relationship between behavioural consistency and need strength. Some control for need strength or trait extremity is essential for reasons developed below. It is difficult to conceive of anyone knowingly endorsing all the items on a scale with consistent and extreme ratings (e.g., "1" or "7") and admitting to a high degree of behavioural variability. Classical conceptions of traits sound much like modern notions of behavioural consistency and have referred to behavioural regularity in such definitions as "the consistency of an individual's response to a variety of situations" (Pervin, 1970, p. 57), and "organized dispositions within the individual which are assumed to have some generality in their manifestations across a variety of stimulus situations" (Wiggins, 1973, p. 320). These definitions imply an inverse relationship between cross-situational variability and trait extremity such that those individuals close to either end of a bipolar trait dimension will exhibit behavioural manifestations of the trait (or its opposite) with greater regularity across diverse contexts, and hence greater consistency, than those who fall at the midpoint on the same dimension.

Stones and Burt (1978) have viewed the association between consistency and extremity from a purely statistical perspective. Presenting some supporting evidence, they surmised that it would be "improbable that a highly

polarized (i.e., extreme) central tendency score could be achieved, except where dispersion was relatively low" (p. 383). In fact, it would not only be improbable but impossible. Ceiling and floor effects of bounded scales set varying upper limits on the variances of sets of points according to their means along the scale. Figure 1 shows the maximum variances of sets of values with given means along a bounded and continuous scale ranging from 0 to 1. The choice of range of the scale is arbitrary. For example, as the proportion of item endorsements on a personality questionnaire becomes extreme, the respondent's trait scale value concomitantly becomes extreme (approaches 0 or 1 in Figure 1) and the consistency of the item responses increases. If such a set of responses to a personality rating questionnaire is used to establish simultaneously the respondent's location on the trait dimension and level of consistency, as is the case with one of Bem and Allen's (1974) manipulations, the function described in Figure 1 will define the relationship between the two individual difference variables. If, on the other hand, estimates of personality and consistency are calculated from distinct data, such as self ratings of personality and self ratings of variability, then the function can be used as a model to predict the expected form of the relationship.

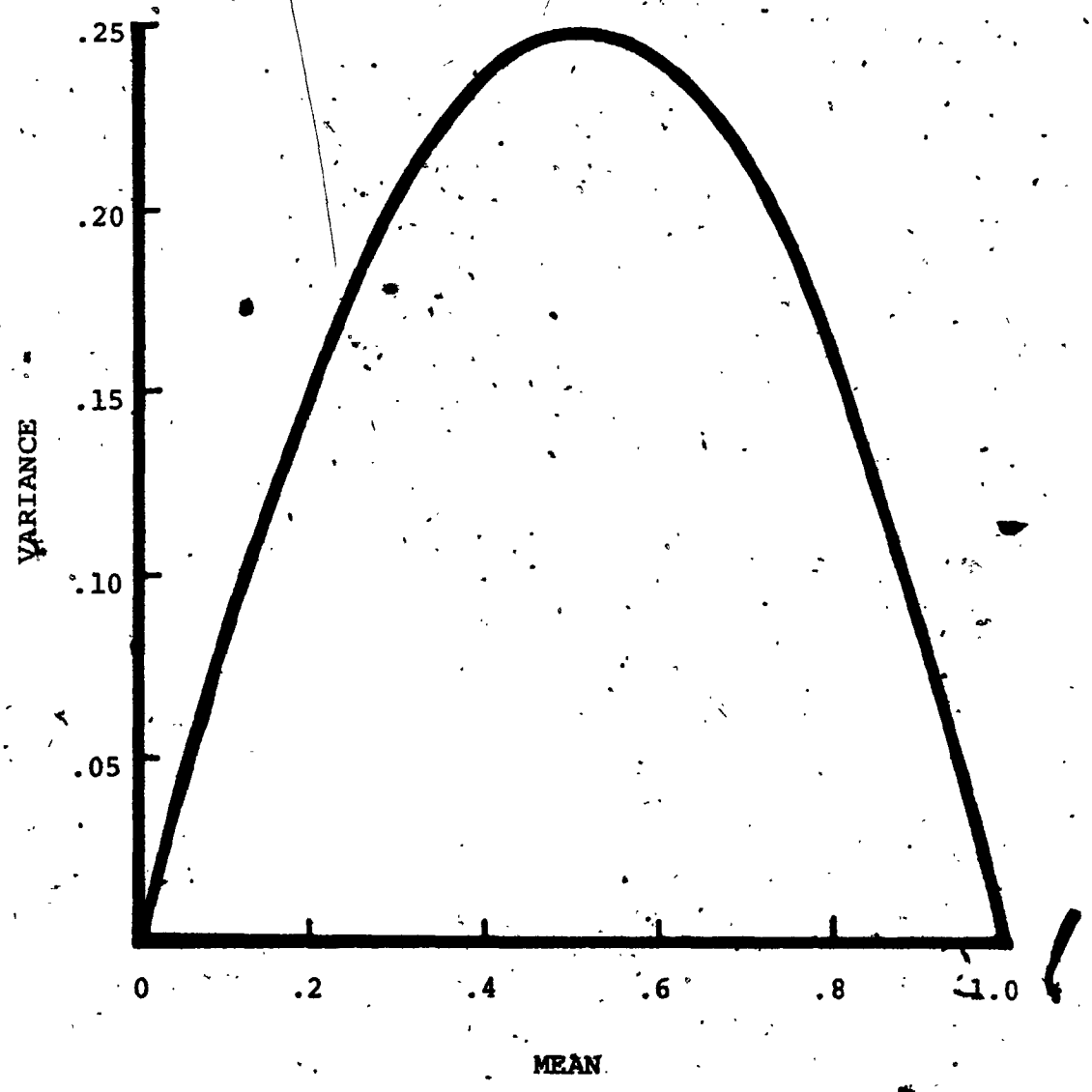

Figure 1. Maximum variances of sets of values with given means on an arbitrary and bounded continuous scale.

Failure to account for a dependency between consistency and personality could lead to serious problems in hypothesis testing. Sorrentino and Short (1977) have documented numerous examples of irregularities in the behaviour of experimental groups of subjects who obtain moderate scores on various motive measures. Discontinuities in the performance of moderate groups in relation to low and high groups in several published studies have prompted the authors to consider a potential relationship between cross-situational consistency and needs. They speculate "it is possible that the moderates are those most inconsistent" (p. 483). This relationship may lead to other problems that are specific to the search for moderator variables. These are described in a later section.

Although Bem and Allen did not report the regression of self rated cross-situational consistency on position on the trait scales, they do present other analyses. Nonsignificant F statistics were found based on a computed index of variability for the lowest, middle, and highest thirds of the total trait score distributions. This variability index was calculated for each respondent as the variance of his/her standard scores on a number of measures of personality. Such a test, however, has problems associated with it in the present instance: (a) Bem and Allen have assumed that differences in personality scores are due to subject variability rather than a lack of test

validity. (Four of the personality scores entering the variance computations were derived from independent ratings on the same items by self, peer, mother, and father. The assumption underlying this analysis is that rater discrepancies merely reflect the cross-situational variability of subject behaviours.); and (b) the tertile split entails some loss of individual difference information when assigning subjects to groups. More appropriate would have been nonlinear regressions of the measures of variability (including self reported variability) on each of the measures of personality. According to the reasoning presented above the relationship between behavioural consistency and total scale scores should be curvilinear with bipolar traits such that consistency would be greatest at the extremes of item endorsement.

Bem and Allen (1974) were able to separate predictable from unpredictable subgroups (matched with respect to trait scores) based on the individuals' subjective decisions concerning behavioural consistency for the dimension of friendliness but not for that of conscientiousness. A "statistical" estimate of consistency was required to differentiate between predictable and unpredictable subjects on the latter trait scale. This index was computed by dividing an individual's response variance across the 23-item Conscientiousness scale by the variance across <u>all</u> responses for the total 86-item multiscale

questionnaire. Bem and Allen claim that this "ipsatized variance" ratio "assumes a value of zero if the individual responds identically to each item on the trait scale and a value of one if he does not 'cluster' the items on a trait scale at all" (p. 515). In actual fact, the ratio assumes a value of zero only when the 23 Conscientiousness items receive identical ratings on the 7-point scale (zero variance) and there is some variance among the remaining 63 items. Otherwise the variance ratio is indeterminate (i.e., zero divided by zero). Furthermore, the maximum value for the statistic is implied to be 1.0, whereas if one computes the maximum variance possible for the 23 item responses (8.98) and divides this by the minimum total variance (4.02) the ratio has an upper bound, in this case, of approximately 2.24.

Unfortunately, it has not been observed that this index of consistency confounds variability of responses to items within a scale with the variability of responses across scales. Consider a situation where two subjects endorse the 23 questionnaire items related to conscientiousness with the same responses using the extreme end points of the rating scales to produce a common variance of 8.98 for the numerator of their ipsatized variance scores. Suppose that one of the subjects is extremely gregarious and endorses the 24 Friendliness items with consistent ratings of "7" while the other is rather withdrawn and records ratings of "1" on the 7-point scales.

Suppose, in addition, that they are both low on the rest of the dimensions measured by the 86-item multiscale inventory and circle the "1" response for the remaining 39 items. Although their response consistency is identical on all of the traits assessed by the questionnaire, the ipsatized variance scores for Conscientiousness would be 1.01 for the first individual (8.89/8.76) and 2.06 for the other (8.89/4.32). The cause of this difference would be their differing levels of friendliness.

The meaning of these anomalies with respect to behavioural variability on the dimension of conscientiousness is unclear. What is clear is that Bem and Allen have allowed information from response consistency on potentially "irrelevant" dimensions (i.e., in their view, idiographically not characteristic of the subject) to determine an estimate of the consistency of behaviours associated with a specified dimension. As illustrated above, two individuals could have identical responses to the 23 conscientiousness items and yet their estimates of consistency on this dimension could be different simply by having different responses to the remaining (heterogeneous) items. This is certainly not in keeping with the central argument punctuating the Bem and Allen paper which maintains (following Allport) that "individuals differ not only in the ways in which traits are related to one another in each person but . . . they differ also in terms of which traits are even relevant"

(p. 509; see also Turner & Gilliland, 1979).

In summary, Bem and Allen (1974) have: (a) argued that individual differences in behavioural consistency would moderate the relationship between various assessments of personality such that the more consistent individuals would yield greater predictive validity coefficients; (b) supported their argument using the trait of friendliness by separating subjects into self-admitted-consistent (predictable) and variable (unpredictable) subgroups; (c) further supported their hypothesis using the trait of conscientiousness but only after devising a consistency measure based on individual item response variability; and (d) recognized and attempted to control for the possibility of a dependency between trait scores and consistency scores, although they report evidence of no relationship using a third index of behavioural consistency.

The Kenrick and Stringfield Study

In not considering the empirical relationship between ratings of behavioural consistency and trait extremity or position on a bipolar trait dimension, data purportedly bearing on differences in predictor-criterion correlation coefficients for consistent versus inconsistent groups may be spurious and irrelevant. Kenrick and Stringfield (1980) sought to replicate and extend Bem and Allen's (1974) findings to a greater number of traits, arguing that each person will be consistent on some dimension(s) of behaviour. Ratings of personality and subjective estimates

of consistency were obtained on 16 one-item bipolar trait scales by self, a parent, and a peer. In addition, each person specified from a total of 16 traits the one most consistent and the one least consistent dimension for the target. The greatest discrepancies in the convergence of self-other personality estimates were in comparing ratings on the subjects' most consistent dimensions with ratings on the least consistent. For example, when correlating the subjects' self report scores on what each identified as his/her "most consistent" dimension, regardless of which dimension, with peer ratings on the same scales, the validity index was a quite substantial .61, computed across all subject-peer dyads. For those traits that they identified as their "least consistent," the correlation was only .12 (Table 2, p. 95). Additionally, in a trait-by-trait replication of Bem and Allen's procedure, Kenrick and Stringfield classified subjects as consistent or inconsistent on each trait dimension separately. Greater self-peer validity correlations were noted for 11 out of 16 traits for the groups consisting of subjects above the median on self rated consistency (Table 6, p. 97).

Rushton, Jackson, and Paunonen (1981) have argued that the Kenrick and Stringfield (1980) study is seriously deficient in several respects. First of all, single-item scales provide rather poor measures of complex human attributes (Nunnally, 1978, p. 210) due to their inherent

unreliability. Multi-item measures are preferred in establishing a respondent's location along some dimension of behaviour. Second, the "variables" being correlated in their primary analysis are enigmatic since these self-other correlations are based on different subjects and different traits. (Not all subjects were most consistent, or least consistent, on the same dimensions.) The Kenrick and Stringfield coefficients are not based on the same trait measured across different individuals nor different traits measured across the same individual but different traits and different individuals correlated across accompanying criterion ratings. Differentially skewed trait score distributions, under these conditions, could yield artificially high correlations (as observed by Kenrick & Braver, 1982, p. 183).

The third and most important problem with the study is that the authors, unlike Bem and Allen (1974), failed to account for the hypothesized relationship between estimates of consistency and trait position. If the two were mutually dependent, separating consistent from inconsistent respondents would be, concurrently, differentiating between those whose trait scores fall at the extremes of the bipolar personality dimensions and those who assume an intermediate position, respectively. A hypothetical situation is illustrated in Figure 2. Any correlation using the scores on this personality variable will be inflated for the "consistent" group since predominately

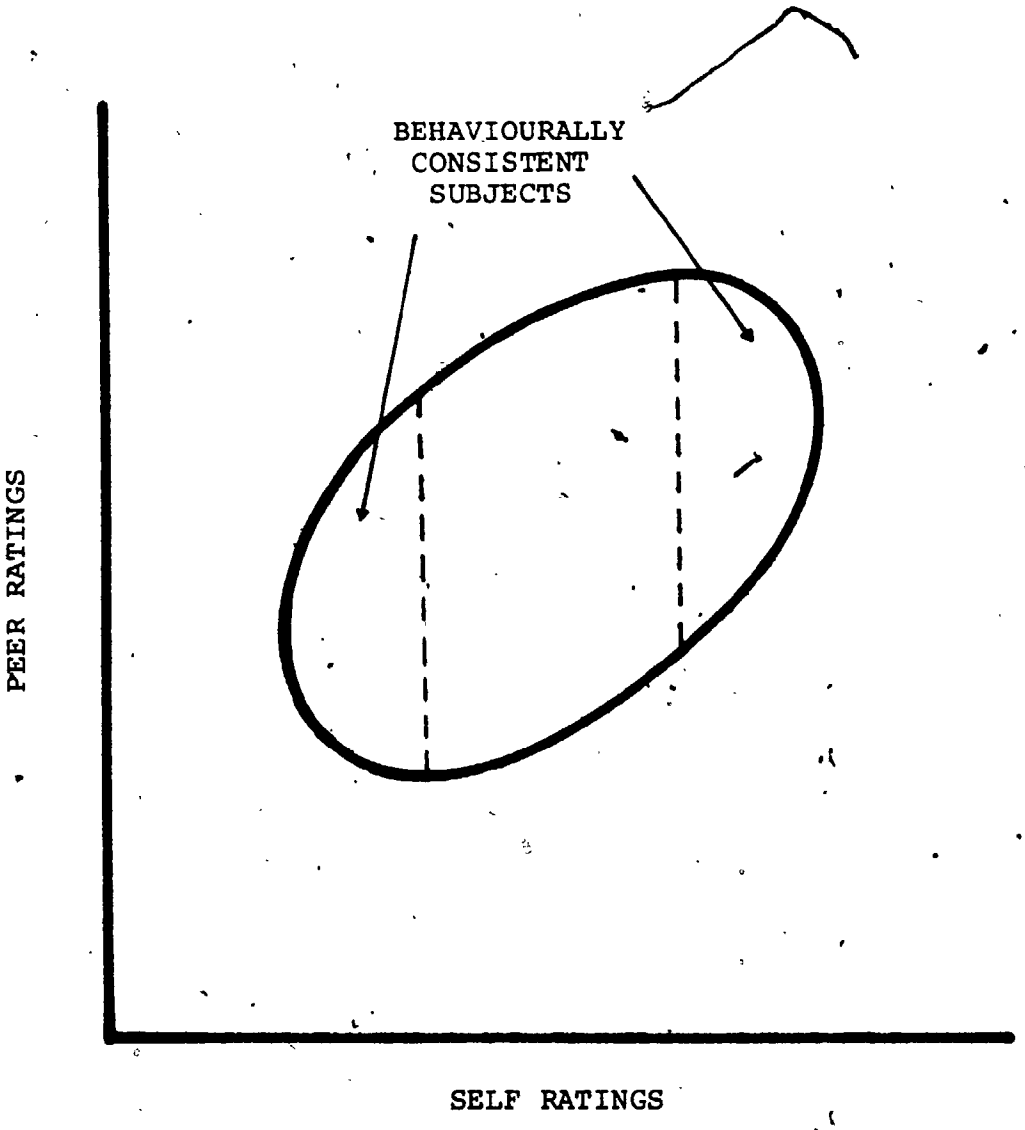

Figure 2. Scatterplot illustrating the relationship between self ratings and peer ratings on a hypothetical trait dimension.

intermediate scale scores will have been removed. Conversely, such correlations for the "inconsistent" group will necessarily suffer attenuation resulting from a restriction of variable range. The differences in predictive validities for the two groups becomes exaggerated. This potential problem also exists in an experimental study by Turner (1978, Table 3) and an investigation reported by Schutz (1979, Table 8).

The Kenrick and Braver Study

Kenrick and Braver (1982) have responded to the above criticisms by Rushton et al. (1981) of Kenrick and Stringfield's (1980) study by presenting new data and "corrected" analyses. They conceded that their primary analysis based on correlations across different subjects and different traits was inappropriate and, as mentioned earlier, could yield misleading results. Their modified analyses consider trait dimensions individually; a validity coefficient was computed on just those subjects who identified a trait as their "most consistent" and was compared with the validity computed across all subjects for that trait. The average "idiographic validity" for the consistent subjects, across all 16 traits, was .47, whereas the average "nomothetic validity" was only .24.

An obvious problem that occurs with Kenrick and Braver's "corrected" analysis is that the sample sizes differ for the "idiographic" and "nomothetic" correlations. In the latter case, all subjects determine the validity

coefficients. The idiographic validities, however, are based on only that subset of individuals who claim that the trait is their most consistent. For example, suppose that Kenrick and Braver's sample consisted of 160 individuals (approximately their actual sample size) measured on the 16 dimensions of personality. Suppose further that there is no differential preference among the subjects over which dimension they consider their behaviours to show the greatest consistency. This would mean an expected 10 individuals endorsing each dimension as their "most consistent." The nomothetic validity correlation for any particular trait would be based on an n of 160 and a value of .24 would be highly significant ($p<.001$). The idiographic validity for that trait would be based on only 10 observations and even a correlation of .47 would not be statistically different from zero ($p>.10$). Unfortunately, Kenrick and Braver have not presented the critical reader with the data needed to evaluate fully their "corrected" approach to the issue.

As for the potential problem of inflation/restriction of range effects mentioned earlier, Kenrick and Braver have found this to be a concern for only some of the trait dimensions incorporated in their study. Comparing the trait score variances for their "most consistent" groups with the variances when all subjects are considered, they found 7 out of 16 traits showed inflated values for the selected subsamples. Since the variances of the consistent

groups were actually smaller for the remaining highly skewed trait score distributions (few subjects being observed at one of the bipolar extremes) and the average variance over all traits for the subgroups was equivalent to the average variance when all subjects entered the computation, Kenrick and Braver concluded that the restriction of range problem "in no way can account for the considerable gain in predictability of the corrected idiographic (.47) over the nomothetic (.24) analysis" (p. 184).

Mention has already been made of problems in comparing the absolute values of the idiographic and nomothetic correlations of Kenrick and Braver due to potentially large discrepancies in sample sizes, with corresponding discrepancies in statistical power. It is therefore difficult to invest much confidence in their conclusions concerning the differential variance problem based on the same imbalanced samples. More importantly, Kenrick and Braver (a) failed to examine the variance issue in the original Kenrick and Stringfield data (1980, Table 6, p. 97) toward which Rushton et al.'s comments were specifically addressed, (b) did not simply apply Bem and Allen's (1974) approach to selecting consistent and inconsistent subsamples having equivalent trait score means and variances, thereby making the variance issue irrelevant while determining if the results of Bem and Allen are replicable, and (c) have changed the focus from a

comparison of behaviourally consistent with variable individuals (e.g., Bem & Allen, 1974; Kenrick & Stringfield, 1980) to a comparison of consistent individuals with the entire sample of subjects--consistent, inconsistent, or otherwise.

Estimators of Behavioural Consistency

The most common measure of behavioural consistency has been to record global, subjective decisions concerning the variability of construct-related behavioural manifestations across situations (e.g., Bem & Allen, 1974; Cheek, 1982; Kenrick & Braver, 1982; Kenrick & Stringfield, 1980; Schutz, 1979; Stones & Burt, 1978; Zanna, Olson, & Fazio, 1980). The next most prevalent has been some index based on the observed variability of responses to questionnaire items, such as the ipsatized variance ratio described earlier (e.g., Bem & Allen, 1974; Campus, 1974; Schneiderman, 1980; Tellegen, Kamp, & Watson, 1982; Underwood & Moore, 1981). Although the two procedures reflect distinct operations, they purportedly measure the same aspect of behaviour. Their convergence, however, has yet to be determined.

The primary intent of the Stones and Burt (1978) study was to evaluate the association of estimates of consistency and trait position. Pursuant to this analysis, a distinction was drawn between subjectively "intuitive" and statistically "formal" methods of estimation. For example, asking subjects how variable they are in performing

trait-related behaviours would be an intuitive estimate of consistency, whereas the variance calculated for a series of item responses would be a formal estimate. The authors found that intuitive but not formal estimates of consistency were, in general, positively related to degree of polarization on two separate profiles of traits. Formal estimates of behavioural consistency were obtained by having subjects "assess the frequency of emission of their behaviors, relevant to different ordinal points along a series of dimensions" (p. 385). Each was asked to distribute 100 points over the divisions of each of several 9-point trait adjective rating scales. The standard deviation of each such distribution was considered the formal estimate of consistency for that trait dimension.

Schneiderman (1980) has found two distinct formal estimates of behavioural consistency to be somewhat related. One of the methods was the variance of a subject's item responses on a personality scale of Friendliness. The second was the subject's variabilities within each of three expressive behaviours related to outgoingness, measured across three experimental role playing conditions. Significant correlations ranging from .30 to .59 between a measure of consistency based on questionnaire responses and one based on more construct-representative behaviours were interpreted as supporting the practicability of a highly formal and objective measure of variability not based on self report.

Turner and Gilliam (1979) have described two further procedures for assessing behavioural consistency using information from self descriptions. One is the "it depends on the situation" alternative on bipolar adjective checklists of the form described by Nisbett, Caputo, Legant, and Maracek (1973) and used by Turner (1978) as a moderator of the predictive validity of personality measures. The other is the failure of a person to write a trait adjective (or its opposite) on a free-recall list of self descriptors, following a report that such omissions were related to behavioural inconsistencies in experimental situations designed to elicit assertiveness (Turner & Gilliland, 1979). The authors concluded that there was little evidence of an association between the two measures of consistency.

Goals of the Present Investigation

In view of the problems and ambiguities identified with principal studies championing individual differences in behavioural consistency as a moderator of the predictive validity of personality assessments, the course of the present research was directed toward the investigation of several issues. A peer rating study of personality was designed to determine the following:

1. The extent to which different attempts to assess the variable of behavioural consistency agree. Proper psychometric procedures are basic requirements for the evaluation of any instrument constructed to scale

individual differences.

2. The extent to which there is a predictable relationship between measures of trait consistency and personality test scores. A logical analysis would suggest that trait-related behaviours become more consistent at the extremes of the trait score continuum.

3. Potential statistical artifacts arising from the idiographic classification of individuals into consistent and variable subgroups when the hypothesized association between consistency scores and need strength is not accounted for (e.g., Kenrick & Stringfield, 1980). Differential trait score variances might result in differential validity coefficients and assumed moderator effects that misrepresent the data.

4. Changes to apparent moderator effects when consistent and variable subgroups are formed having equivalent trait score means and variances. If trait score dispersion differences between the groups have contributed to discrepancies in predictive validity, the elimination of these differences should redress some of the observed imbalance in predictability.

5. Evidence for the moderating effect of behavioural consistency when the technique of moderated multiple regression is applied to the data. This procedure, customarily overlooked, does not require the classification of subjects into distinct subgroups and is the preferred analytic method in the search for moderator variables

(Saunders, 1956; Tellegen et al., 1982). Its details are outlined in Chapter 2.

6. Judgmental uncertainty as a possible moderator of the validity of personality ratings. It is not unreasonable to expect that trait attributions by raters will be accurate as a function of the degree of confidence expressed in making decisions about target behaviour probabilities. Uncertainty may reflect random or "best guess" strategies that yield inferior indices of judgmental validity.

Secondary analyses of the personality rating data were designed to examine other issues that have been used to impugn traditional nomothetic assessment formulations. Specifically explored were

7. Increments in predictive validity as a function of the aggregation of measurements of personality. Much of the concern surrounding contemporary assessment strategies has been due to the ubiquity of mediocre coefficients of validity. Such coefficients, however, frequently have referred to one-item measures of personality. Based on the psychometric theory of measurement error and the reliability of behaviour samples (e.g., Nunnally, 1978) as well as the results of recent studies (e.g., Epstein, 1979, 1980; Rushton et al., 1981), it was anticipated that the validity of behaviour predictions would be greater for multi-item as opposed to single item measures of the same construct.

8. The similarities of factor structures underlying ratings of personality by targets and peer judges using published multi-item scales and ad hoc single item measures. Other studies have shown peer perceptions of the organization of target personalities to correspond closely to the structure determined through target self ratings (e.g., Jackson & Guthrie, 1968).

9. The factorial validities of ratings of personality. The observation of a similarity between factor structures derived from target personality ratings by self, peer acquaintances, and by strangers has been used to imply that conventional peer ratings are not valid since the perceived organization of target personalities is invariant with changing degrees of target-rater acquaintanceship (D'Andrade, 1964; Mulaik, 1964; Passini & Norman, 1966; Shweder, 1975). This inference is unquestionably a non sequitur. A correspondence between self and other determined factor structures is a necessary but not sufficient condition for valid peer ratings. The sufficiency condition is satisfied by observing an adjacency of perceived and actual target locations in the factor space.

CHAPTER 2

MODERATED MULTIPLE REGRESSION

Most of the research discussed in the preceding introduction employed a methodology for the study of moderator variables that classified subjects as either predictable or unpredictable in their behaviours. This methodology involves identifying predictable subjects as those who obtain scores above (or below) an arbitrary cutoff point (usually the median) on some variable potentially moderating the predictive efficacy of different assessments of behaviour. Alternatively, the moderator variable could form a natural dichotomy of supposedly predictable/not predictable groupings as with sex as a moderator of the relationship of various predictors with academic achievement (e.g., Abelson, 1952). In either case, the moderator variable is treated as dichotomous and individual differences within categories of membership are not considered in subsequent analyses.

A more general approach to the investigation of moderating effects, which provides for moderators as continuous variables, was initially proposed by Saunders (1956). The procedure of forced dichotomies (e.g., Bem & Allen, 1974) assumes that linear regression lines in the prediction of criterion behaviours have different slopes for high and low moderator groups. The generalization discussed by Saunders, however, assumes that an infinity of

regression lines vary in slope as a continuous function of the moderator variable. This function can define an inverse relationship, such that increasing levels of the moderator are systematically associated with decreasing slopes, or a direct relationship. The advantage, of course, in such a treatment of moderators is that it is entirely consistent with psychological theory; few personological variables are thought to be inherently dichotomous. Furthermore, all subjects are included in the regression analysis rather than subsets employed in separate regressions, resulting in greater power of statistical tests of hypotheses concerning regression parameters.

Moderated multiple regression is a procedure that has as its basis a straightforward modification of the general multiple regression equation (Saunders, 1956; Wiggins, 1973). Consider the model of linear multiple regression in the two-predictor case:

$$\underline{Y} = \beta_0 + \beta_1\underline{X} + \beta_2\underline{M} + \varepsilon \ .$$

The variable $\underline{Y}$ is conceived of as a weighted linear combination of the predictors $\underline{X}$ and $\underline{M}$, a regression constant, and an error term. In this example, the units of observation can be represented as points in a three-dimensional Euclidian space with coordinates $(\underline{y},\underline{x},\underline{m})$. The coefficients for the equation are chosen to fit a flat

plane somewhere in this space so that the sum of the squared distances (measured parallel to the y-axis) of the points in the space to the fitted plane is at a minimum. Such a plane, also known as a response surface, represents predicted criterion scores and an example is illustrated in Figure 3. The data points are contrived and not depicted in this illustration.

The model underlying moderated multiple regression has as its foundation the linear multiple regression equation above with the addition of a multiplicative term reflecting the product of the predictors:

$$\underline{Y} = \beta_0 + \beta_1\underline{X} + \beta_2\underline{M} + \beta_3\underline{XM} + \varepsilon .$$

The product term is analogous to the two-way interaction term of the general linear model underlying analysis of variance formulations (Saunders, 1956) and has been referred to as bilinear in form (Green & Carroll, 1976). It accounts for the joint or conditional effects of predictors on criterion (Cohen & Cohen, 1975); the relationship between the criterion and a predictor, as with the dependent variable and independent variable in analysis of variance, is conditional on the status of a third variable. This third variable is the moderator variable and the interaction term accounts for the moderator effect. Thus, in the present context, the relationship between two different assessments of personality would be conditional

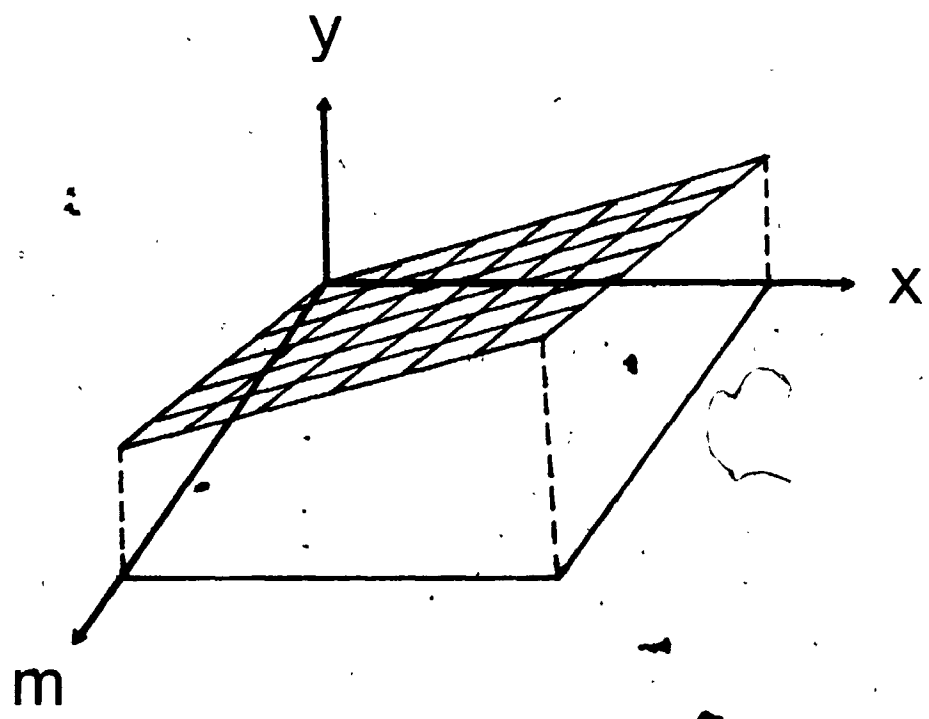

Figure 3. Hypothetical multiple regression of Y on X and M.

on the level of behavioural consistency. It should be mentioned at this point that, as M moderates the relationship between X and Y, X similarly moderates the relationship between M and Y. Which predictor is described as the moderator follows from theory and is discussed more fully below. Furthermore, the model is readily generalized to situations with more than two predictors, having several two-way interaction components (Saunders, 1956) and n-way interaction terms (Cohen & Cohen, 1975).

The geometric effect of including the product term in the multiple regression equation is to allow for the bending or distortion of the response surface fitted to the observed data. A hypothetical illustration is presented in Figure 4. The light lines on the ruled surface are parallel to the XY and MY planes and are straight. These lines are a subset of an infinity of regression lines that vary in slope across the response surface as a function of the predictors. For example, the regression of Y on X is positive at small values of M. As M increases in value, the regression of Y on X passes through zero and becomes increasingly negative. If M is a measure of behavioural variability and X and Y are different assessments of personality, the relevance of the regression model to current issues is apparent; the validity of the assessments would be expected to vary as a function of behavioural consistency. The median split approach to the analysis of moderator variables essentially bisects the

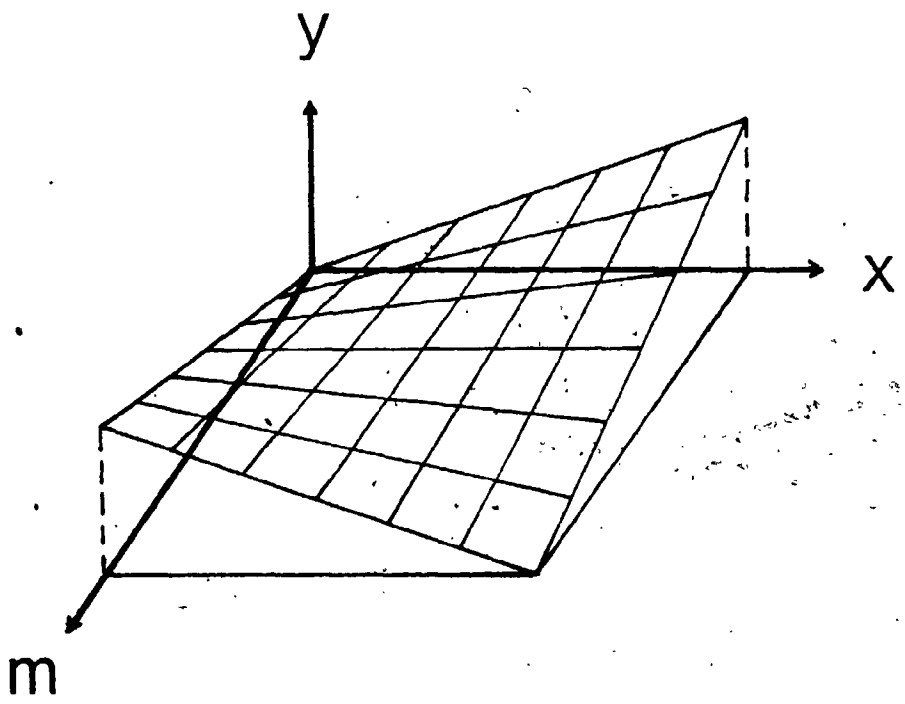

Figure 4. Hypothetical moderated multiple regression of Y on X, M, and XM: Interaction term with negative regression weight. (After Saunders, 1956.)

three-dimensional space in a plane parallel to XY at a point somewhere along M (i.e., the median of behavioural consistency or variability). The data are collapsed across M within the high and low M groups and the regression of Y on X examined separately for the two.

The interaction term for the example of Figure 4 has a negative regression coefficient since conjointly high values of the predictors (hence, a large product term) yield lower values of the criterion than when either predictor is large and the other small. Figure 5 illustrates the case when the XM product has a positive weighting in the equation. Small values of the moderator are associated with negative regressions of Y on X whereas positive slopes are evident for larger values. This would occur in the present context if (a) the moderator was scored in the direction of behavioural consistency rather than variability and the moderating effect was as hypothesized, or (b) if behavioural variability was being measured and the moderating effect was such that the more variable subjects were in fact more predictable in their behaviours.

A sequential or Hierarchical procedure is used to apply the moderated multiple regression model to a set of observed data. The first step is to regress the criterion Y on predictor X and evaluate the variance common to the two variables. The moderator variable is then included in the equation and its linear contribution towards the

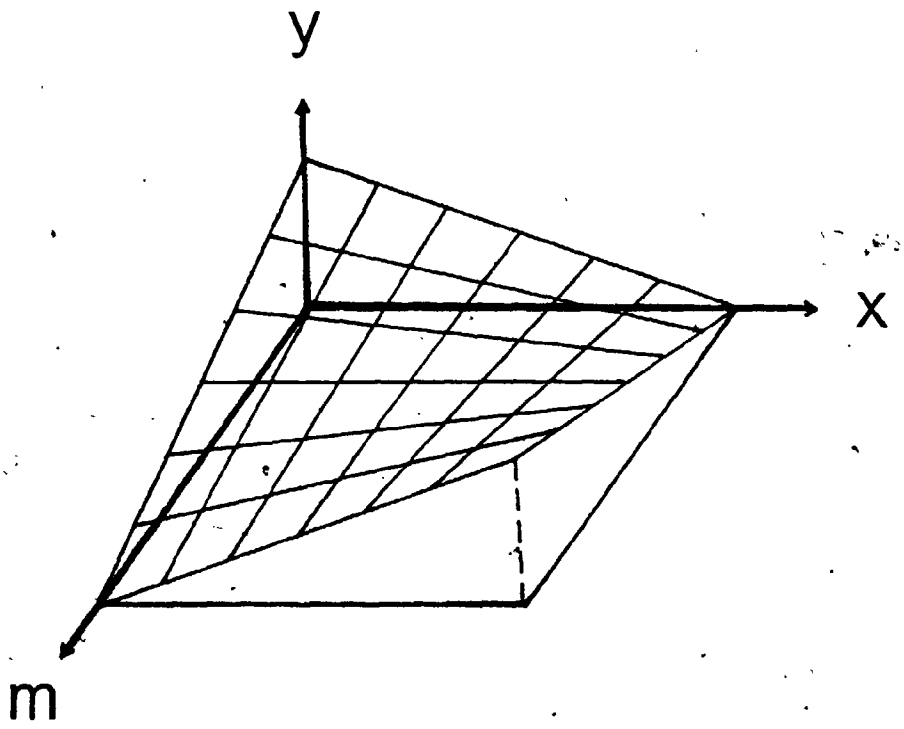

Figure 5. Hypothetical moderated multiple regression of Y on X, M, and XM: Interaction term with positive regression weight.

prediction of Y evaluated. At this point a flat "least-squares" plane has been fitted to the data and the errors in prediction are reflected in the distances of the observed data points from their respective predicted values on the response surface. The XM product is finally entered into the prediction equation and the response surface is "bent." If the variable M moderates the regression of Y on X, then the bending of the surface will minimize, nontrivially, errors in prediction and the multiple correlation of predictors with criterion will improve significantly.

An hypothesized moderating effect of a variable on the relationship between two other variables will be called into question if either (a) the multiplicative component adds little to the explanation of residual criterion variance after partialling the linear contributions of X and M (that is, the response surface remains essentially flat and the regression of Y on X is constant at all levels of M), or (b) the sign of the regression coefficient for the interaction term is not consistent with expectations according to psychological theory. The statistical significance of any term added sequentially to a prediction equation already containing k terms can be evaluated (after Cohen, 1968) by the ratio

$$F = \frac{R^2_{Y\cdot 1\ldots k,k+1} - R^2_{Y\cdot 1\ldots k}}{(1-R^2_{Y\cdot 1\ldots k,k+1})/(n-k-2)},$$

where n is the sample size, and with degrees of freedom equal to 1 and $n-k-2$. The signs of the regression coefficients for the product terms are expected to be negative in the present investigation if the moderator is scored in the direction of behavioural variability and positive if high scores on the moderator reflect behavioural consistency.

CHAPTER 3

METHOD

Subjects

A peer rating study was carried out on a sample of 45 roommate pairs solicited from an undergraduate student residence at The University of Western Ontario. A stipend was paid to each volunteer at the end of the testing sessions. The sample ranged in age from 18 to 22 years and consisted of 10 male and 35 female pairs. The study was carried out within the second last month of the academic year, ensuring the subjects had had ample opportunity to become acquainted with their respective roommates.

Procedure

A recruitment roster for a "Study of Personality" was placed in the lobby of a co-ed residence requesting roommate pairs as volunteers for the study. The paid volunteers were administered rating tasks in two sessions completed on two separate days. During the first session, the 90 subjects were asked to complete a battery of paper and pencil self rating personality questionnaires. They were assured, by means of coded forms, of the anonymity of their responses. The same battery of materials was administered for the second session, when subjects were presented with the challenge of making accurate judgments of the behaviours of their respective roommates. Except for identifying the target of the ratings, all instructions and forms were identical for the self and peer ratings.

All rating forms were examined for missing or improper responses before subjects left the testing room.

Materials

The 352 behaviour-situation type items of the Personality Research Form (PRF Form-E; Jackson, 1974) were used as the primary measure of personality. The construct validity of the PRF has been established largely through demonstrations of its predictive utility with respect to various behavioural criteria. Such criteria have included the marital adjustment of young couples (Meyer & Pepper, 1977), objective measures of social interactions (Gifford, 1981, 1982), employment preferences (Rothstein & Jackson, 1980), and a variety of types of behaviour ratings made by close acquaintances (e.g., Harris, 1980; Jackson, 1974). This inventory consists of 16 items for each of 20 content scales and 2 validity scales. Self and peer judgments of personality were made with respect to probabilities of behaviour recorded on 9-point rating scales (Appendix A). For example, subjects were asked to consider the statement "I feel very sorry for lonely people" and then choose a number on a 1-9 rating scale where a "1" response represented a decision of the behaviour being "extremely uncharacteristic" of the person being rated, a "9" indicating an "extremely characteristic" behaviour, and a "5" representing a behaviour that was judged neither characteristic nor uncharacteristic of the target ("neutral").

The second battery of ratings consisted of a 20 item set of judgments using trait defining adjectives that were chosen to exemplify both poles of each of the 20 PRF dimensions of personality (Appendix B). To illustrate, one of the bipolar dimensions assessed by the inventory is labelled Affiliation. Subjects first read a brief description of a person who would typify the characteristics of a high scoring respondent on the Affiliation scale--"enjoys being with friends and people in general; accepts people readily; makes efforts to win friendships and maintain personal ties." Below this was a 9-point rating scale that had as its end-point anchors "sociable" and "withdrawn." Judges were instructed to rate the target (self or peer) on "sociable" being "extremely characteristic" ("9") to "withdrawn" being "extremely characteristic" ("1"), with "5" as "neutral". The adjective anchors and prose descriptions were chosen in consultation with the PRF manual (Jackson, 1974, pp.6-7) and the test's author, in an attempt to capture the essence of both poles of each of the 20 dimensions of personality.

In addition to each rating of target need strength, judges were asked to estimate the variability of target behaviours with respect to each of the 20 dimensions and their degree of confidence in making the trait rating (Appendix B). After responding to the question "How characteristic is SOCIABLE or WITHDRAWN" of the person being rated (self or peer), they were asked to state "how

confident" they were in that decision (a "9" indicating "extremely confident" to "1" indicating "not at all confident" with "5" as "average") and "To what extent do behaviours related to the trait dimension above vary from situation to situation" for the target ("9" being "extremely variable", "1" being "extremely consistent", and "5" as "average").

Prior to the administration of the roommate peer rating forms, each judge was asked to complete a 7-point degree of acquaintance rating (Appendix A).

CHAPTER 4

RESULTS

Analyses of the data were designed primarily to evaluate the psychometric properties of various indices of behavioural consistency that have been commonly used in idiographically-based studies of personality and to determine the generalizability and robustness of previous findings related to the predictability of behaviour as moderated by individual differences in trait consistency. The confidence expressed by a rater in making a trait attribution was also evaluated as a moderator of the convergence between different assessments of personality.

Secondary analyses were concerned with comparing the relative magnitude of rating validities based on ad hoc, single item adjective scales with those that emerge when personality is assessed by established multi-item scales of known properties. Additionally, the factor structures of self and peer ratings were determined and examined for convergence, both of factor composition and of self and peer estimated target locations in the factor space.

Prior to these analyses, the data were checked for missing or random responses. Of the PRF item responses, 27 were left blank on the self rating forms (.00085%) and 37 were blank for peer ratings (.0012%). These missing values were replaced with a "5" reflecting the centre or neutral point of the 9-point bipolar rating scale. No other blanks

were found in the remaining data. Tendencies towards randomness in responding to the questionnaires were inferred from the elevation of the PRF Infrequency scale. This collection of items with extremely low probabilities of endorsement is used to detect scoring errors, careless responding, and low-level linguistic abilities (Jackson, 1974). It was decided that none of the respondents in the present study demonstrated Infrequency scores sufficiently elevated to cause his/her removal from the data pool.

Because of the relatively small number of males in the sample (20 out of 90), all analyses at the group level were carried out without regard to subject gender. Furthermore, the observed high degree of acquaintanceship between all members of the roommate pairs with very little variability (mean 7-point acquaintanceship rating = 5.8, variance = 1.01) precluded the use of this variable in the analyses that follow.

Evaluation of Behavioural Consistency Measures

Overview. To be considered seriously as a practicable and important variable for research, the notion of behavioural consistency must be unambiguously defined such that there exists a reasonable consensus with regard to its interpretation and empirical operationalizations. Measures thought to scale individual differences in behavioural consistency must demonstrate some degree of convergence among themselves and yet be distinguishable from theoretically independent variables.

Most studies in the past have incorporated a simple one item decision, using a 5-point or 7-point rating scale, as an intuitive measure of behavioural consistency or variability (e.g., Bem & Allen, 1974; Kenrick & Stringfield, 1980; Zanna et al., 1980). Bem and Allen (1974) advocated and used a variance ratio as a formal index of behavioural consistency, in addition to the subjective global self rating. This "ipsatized variance ratio" was computed for an individual as the subject's response variance across the items of a particular trait scale divided by the person's response variance across all items in a multiscale test battery. Others have used a nonipsatized variant, computed as the simple variance of a person's item responses for a trait questionnaire (Schneiderman, 1980; Underwood & Moore, 1981).

Traditional procedures for evaluating the validity of personality scales frequently include (a) an assessment of the convergence of self-ratings on the measure with ratings of the self by well-known others, and (b) the determination of convergence among alternate forms of the measure. The following analyses examine these relationships and also determine the mutual dependency between common indices of behavioural variability and personality measurements.

Self-peer rating convergence. The 90 self ratings on the 9-point consistency judgments were correlated with the corresponding roommate ratings of consistency; for each of the 20 PRF adjective trait scales. Column 1 of Table 1

Table 1

Intercorrelations of Indices of Subject Consistency for 20 Trait Dimensions

	a	b	c	d
ABASEMENT	-.23	-.08	-.15	.58
ACHIEVEMENT	.24	-.01	.00	.70
AFFILIATION	.21	.30	.23	.70
AGGRESSION	.02	-.13	-.15	.60
AUTONOMY	.06	.22	.05	.67
CHANGE	-.02	-.04	-.06	.66
COGNITIVE STR.	.28	-.05	-.07	.70
DEFENDENCE	-.03	.19	.09	.72
DOMINANCE	.09	.23	.18	.80
ENDURANCE	.00	.12	-.01	.66
EXHIBITION	.06	.10	.09	.78
HARMAVOIDANCE	.10	.04	.05	.73
IMPULSIVITY	-.11	.13	-.01	.75
NURTURANCE	.13	.20	.13	.70
ORDER	-.03	.03	-.12	.71
PLAY	.04	.07	-.11	.62
SENTIENCE	.06	.03	.08	.62
SOC. RECOGNITION	-.01	.13	.10	.58
SUCCORANCE	.03	.15	.10	.70
UNDERSTANDING	-.03	.09	.07	.61
MEAN	.04	.09	.02	.68

Note. $n=90$, $r>.175$, $p<.05$.

a- Self with peer adjective ratings.
b- Self adjective ratings with ipsatized variance ratios.
c- Self adjective ratings with scale variances.
d- Ipsatized variance ratios with scale variances.

shows these correlations to be quite low with a mean of .04. Schutz (1979, p. 44) has found evidence for greater self-peer rating convergence on similar ad hoc PRF adjective scales, reporting a mean correlation of .27.

Convergence of formal and intuitive estimates. Table 1 column 2 presents the correlations of self rated variability scores with Bem and Allen's variance ratio index of consistency. The ipsatized variance ratio for each respondent was computed for each trait as his/her response variance across the 16 PRF items within a scale divided by the person's response variance across all 352 PRF items. In general, these two measures of the same construct share little variance. A further chi-square analysis with both of these indices involved classifying each subject on each trait dimension as either consistent or variable. Bem and Allen's procedure for orthogonalizing consistency and need strength was followed and each person received two binary (i.e., consistent or variable) scores for each trait dimension; one based on self ratings of variability, and the other based on the ipsatized variance ratio. None of the 20 resultant 2 by 2 contingency tables had cell proportions that differed significantly ($p=.05$) from those expected by random assignment. The simple variance of the PRF item responses for a scale as an index of behavioural consistency also shows essentially no correspondence with subjective estimates of variability, as seen in column 3 of Table 1. As would be expected,

however, the variance ratio does overlap with the simple item variance index since their computation is based on shared data (column 4).

Consistency and trait extremity. Reference was made in the introduction to the similarity between classical definitions of traits or needs and modern conceptions of behavioural consistency. It was suggested that any distinction between the two becomes obfuscated as need strength increases such that a high degree of behavioural variability for a trait would be, by definition, precluded at an extreme level of the trait. Conceptually, one would expect logical decisions about behavioural consistency to be predictably related to decisions about trait extremity: extremely high or low trait scores should engender high consistency ratings, while neutral values on bipolar trait dimensions could imply either high or low behavioural variability. (This relationship was not found on the average across 12 dimensions in a recent study by Cheek, 1982; cf. Rushton et al., 1981.)

To test for the expected nonlinearity between consistency and trait level, eta coefficients were computed between the subjective variability ratings and the experimentally independent trait adjective endorsements, for each trait. Eta, or the correlation ratio, is a measure of the extent to which one variable can be used to predict another by a curvilinear (or straight) line of best-fit (Glass & Stanley, 1970, pp. 150-152). Columns 1

and 2 of Table 2 contrast the product-moment correlational index of linearity with the eta coefficient of nonlinearity. In most cases, substantial increases in predictability result from a curvilinear regression of variability scores on trait scores. Although one might expect the correlational index of linearity to be small in this analysis, skewed trait score distributions would yield increasing degrees of linear association. In this sample, most subjects rated themselves relatively high on Nurturance (Appendix C). The result is a negative linear relationship between Nurturance and variability (a positive relationship with consistency) since few subjects fell in the lower left section of the hypothesized inverted-U plot of Nurturance by variability scores. Relatively low self ratings on Defendence, alternatively, yielded a positive correlation with variability due to a lack of data points in the lower right section of the scatterplot. The relative sizes of these coefficients in column 1 of Table 2 closely mirror those found by Schutz (1979) in his analysis of the same relationship using similar PRF adjective rating scales.

The relationship between trait level and behavioural consistency can also be examined by the product-moment correlation between the variability ratings and an index of trait extremity, as shown by Rushton et al. (1981). For a bipolar dimension of behaviour, trait extremity can be conceived of as the absolute distance of a person's

Table 2

Correlations (a) and Eta Coefficients (b) Between Self Ratings of Consistency and Trait Adjective Endorsements, and Correlations (c) of Consistency with Trait Extremity, for 20 Dimensions of Behaviour

	a	b	c
ABASEMENT	.04	.45	.31
ACHIEVEMENT	.46	.62	.54
AFFILIATION	.49	.61	.49
AGGRESSION	-.34	.63	.54
AUTONOMY	.23	.35	.29
CHANGE	.09	.46	.36
COG. STRUC.	.35	.43	.27
DEFENDENCE	-.44	.63	.52
DOMINANCE	.30	.45	.38
ENDURANCE	.40	.59	.56
EXHIBITION	.15	.36	.30
HARMAVOID.	-.11	.48	.43
IMPULSIVITY	-.11	.56	.52
NURTURANCE	.58	.70	.68
ORDER	.36	.65	.57
PLAY	.25	.45	.41
SENTIENCE	.48	.56	.45
SOC. RECOG.	-.04	.37	.19
SUCCORANCE	-.21	.56	.47
UNDERSTAND.	.54	.58	.55

Note. $n=90$, $r>.175$, $p<.05$.

location towards either pole from some point of neutrality. For the present trait rating data, a "5" on the 1-9 rating scales was identified as the neutral point falling between two opposite trait descriptors. Extremity scores were calculated as the numerical distance between a subjects's self rating on the scale and this middle value. The linear relationship between variability and trait extremity is shown, by traits, in column 3 of Table 2. These values are somewhat higher than those presented by Rushton et al. (1981) in their analysis of the same relationship.

Although the eta coefficients of Table 2 can refer to U-shaped functions, inverted-U functions, or more complex patterns of data points, it is unlikely that the present data are not of the expected quadratic form (i.e., inverted-U). The correlations in column 3 of consistency with extremity would signal a suspicious scatter of points with zero or negative coefficients. Polynomial regressions, however, were carried out to confirm the form of the relationship by fitting quadratic functions to the observed variability by trait level data. For all but three traits, the square of the trait scores added significantly ($p<.05$) to the prediction of variability (Appendix D). All beta coefficients for the second-order polynomial terms were in the expected (inverted-U) direction.

Eta coefficients and product-moment correlations demonstrating the associations of peer rated behavioural variability with trait ratings by peers are shown in Table 3. These data reproduce the general patterning of the data based on self ratings. Fitting quadratic functions to the data yielded significant increases in the prediction of variability with the square of trait scores for all 20 dimensions (Appendix D) substantiating the assertion that the form of the relationship between the two variables can be described as an inverted-U.

Bem and Allen's (1974) variance ratio is not contingent on subjective estimates of consistency but, rather, is a calculated score based on measured response variability. Since the variance ratio is derived from the same item data used to identify a respondent's location on a behavioural dimension, it is of interest to determine the empirical relationship between this index of a target's behavioural consistency for a trait and the target's location on the trait dimension. Table 4 column 1 reports the correlations, across subjects, of the ipsatized variance scores with the mean of the PRF item endorsements by trait. Eta coefficients computed for the same variables are presented in column 2. The computation of eta requires the independent variable (trait scores) to be discrete. Accordingly, the mean 9-point response on the PRF items was rounded to the nearest integer, for each subject and each trait. This column of coefficients demonstrates

Table 3

Correlations (a) and Eta Coefficients (b) Between Peer Ratings of Consistency and Peer Trait Adjective Ratings, and Correlations (c) of Consistency with Trait Extremity, for 20 Dimensions of Behaviour

	a	b	c
ABASEMENT	.12	.29	.21
ACHIEVEMENT	.38	.50	.44
AFFILIATION	.25	.56	.48
AGGRESSION	-.34	.58	.55
AUTONOMY	.24	.65	.61
CHANGE	-.09	.48	.35
COG. STRUC.	.36	.58	.47
DEFENDENCE	-.18	.70	.67
DOMINANCE	.24	.48	.40
ENDURANCE	.15	.42	.37
EXHIBITION	.11	.46	.43
HARMAVOID.	.06	.47	.44
IMPULSIVITY	-.17	.44	.33
NURTURANCE	.36	.57	.52
ORDER	.02	.49	.43
PLAY	.13	.42	.36
SENTIENCE	.35	.57	.49
SOC. RECOG.	.01	.43	.37
SUCCORANCE	.17	.42	.39
UNDERSTAND.	.41	.52	.41

Note. $n=90$, $r>.175$, $p<.05$.

Table 4

Correlations (a) and Eta Coefficients (b) Between Computed Consistency Scores and PRF Trait Scores, for 20 Trait Dimensions

	Ipsatized variance ratio		Scale variance	
	a	b	a	b
ABASEMENT	-.07	.41	-.05	.24
ACHIEVEMENT	.31	.55	.12	.32
AFFILIATION	.34	.59	.17	.39
AGGRESSION	.02	.60	-.05	.28
AUTONOMY	.18	.54	.16	.33
CHANGE	.24	.44	.11	.25
COG. STRUC.	-.31	.50	.27	.39
DEFENDENCE	-.23	.51	-.12	.28
DOMINANCE	.30	.64	.23	.41
ENDURANCE	.21	.60	.13	.37
EXHIBITION	.43	.63	.34	.45
HARMAVOID.	-.16	.57	-.22	.38
IMPULSIVITY	-.18	.49	-.08	.28
NURTURANCE	.50	.60	.33	.42
ORDER	.00	.54	.01	.29
PLAY	.25	.51	.12	.24
SENTIENCE	.35	.41	.25	.37
SOC. RECOG.	.13	.58	.10	.32
SUCCORANCE	-.16	.47	-.15	.23
UNDERSTAND.	.02	.46	.03	.36

Note. $n=90$, $r>.175$, $p<.05$.

empirically the statistical dependency between an objective operationalization of response consistency and the measurement of need strength when both are derived from the same behavioural observations. Polynomial regressions were carried out to confirm the shape of the function relating the ipsatized variance ratio to need strength. In all 20 cases the function was found to be quadratic or an inverted-U in form (Appendix D).

Columns 3 and 4 of Table 4 present the correlations and eta coefficients of the PRF trait scores with the simple variance index of consistency used by Schneiderman (1980) and adopted by Underwood and Moore (1981). Although there is some nonlinear component in most instances, there is substantially less empirical dependency in this analysis than in that involving the ipsatized variance ratio. Polynomial regressions designed to fit quadratic functions to the data showed significant curvilinearity ($p<.05$) for only 8 out of the 20 traits, although all regression coefficients for second-order terms were in the expected direction (Appendix D). Inspection of the scatterplots of item variances by trait scores revealed, in general, little variability in the responses of individuals to items comprising the PRF scales. This is to be expected when one is considering scales that have been constructed to maximize their internal consistency. The relatively high interitem correlations would preclude high levels of variability for the item endorsements within a scale. In

an analogous context, Underwood and Moore (1981) have found "no linear or quadratic relationship between variability and extremity" in their item response data (p. 782).

Consistency as a Moderator of Judgmental Validity

Overview. Bem and Allen (1974) have argued that idiographic considerations can suggest procedures of personality assessment that yield validity coefficients (or, cross-situational stability indices) substantially greater than those obtained from traditional nomothetic measurement approaches. The assessment paradigm conceived by these considerations has the following components: choice of a psychological variable thought to moderate the validity of different assessments of a trait dimension; individual consideration and scaling of each subject on the moderator variable; division of subjects into high and low scorers on the moderator; determination of the validity of different assessments on the trait for the high and low moderator groups.

Bem and Allen's (1974) idiographically-based study, followed by that of Kenrick and Stringfield (1980), used behavioural consistency as the moderator variable to classify individuals as predictable or unpredictable, yielding differential trait validity coefficients for the two subgroups. The analyses below apply the subtly different procedures of (a) Kenrick and Stringfield (1980) and (b) Bem and Allen (1974) to the current data. Evaluated are differences in interpretation resulting from

the two procedures and the replicability and generalizability of the moderating effect of behavioural consistency with respect to the predictive validity of personality assessments. In addition, more rigorous analyses attempt to identify moderator effects in the data using the technique of moderated multiple regression.

The Kenrick and Stringfield procedure. Kenrick and Stringfield (1980) incorporated the median split methodology of Bem and Allen (1974) in a study designed to replicate and extend the trait consistency findings to a wider variety of behavioural dimensions. For each of 16 sets of trait ratings, respondents were divided into consistent and inconsistent subgroups based on their subjective estimates of behavioural variability (7-point global rating for each dimension), following which self-other personality ratings were correlated. In one set of analyses, they found greater predictive validities for the consistent groups on 11 out of 16 trait dimensions (Kenrick & Stringfield, 1980, Table 6).

Unfortunately, the Kenrick and Stringfield procedure was not identical with that of Bem and Allen since the former did not control for the expected confound between consistency and trait extremity. Rushton et al. (1981) have argued that the consistent groups in the Kenrick and Stringfield study may reflect people having extreme trait scores, resulting in spurious correlations due to restriction/inflation of range effects. Kenrick and Braver

(1982) have countered with "corrected" analyses, claiming that the problem is not apparent in their data.

The following tables are based on splitting subjects into consistent and variable subgroups, by trait, without regard to the observed curvilinearity between trait scores and consistency measures (cf. Table 2 and Table 4). Table 5 summarizes the analysis of the adjective rating validities as moderated by the global self estimates of consistency. Column 1 shows the "nomothetic" correlations where all subjects are considered, in juxtaposition with the "idiographic" correlations of columns 2 and 3 based on consistent and variable subjects, respectively. At a glance, the conclusions of Kenrick and Stringfield seem to be supported since the consistent subgroups have higher predictive validities than the variable groups for 15 out of 20 traits. Upon closer inspection, however, three problems become apparent; (a) only four of 15 differences in validity coefficients are significant at or beyond the .05 level, (b) three of the 20 trait dimension (Achievement, Autonomy, Nurturance) are best predicted by neither consistent nor variable subgroups but by considering the entire sample, and (c) selecting differentially consistent subgroups without regard to trait extremity may have resulted in inflated trait score variances for the consistent (extreme) subgroups and restricted variances for the variable (moderate) groups (Rushton et al., 1981; cf. Kenrick & Braver, 1982). To

Table 5

Differences in Peer Rating Validities for Subjects Above and Below the Median on Self Rated Consistency, for 20 Adjective Scales: Consistency and Trait Level Confounded

	All subjects (n=90)	Consistent subjects (n=45)	Variable subjects (n=45)	(difference)
ABASEMENT	.17	.09	.28	-.19
ACHIEVEMENT	.28	.27	.09	.18
AFFILIATION	.48	.51	.34	.17
AGGRESSION	.34	.25	.45	-.20
AUTONOMY	.23	.18	.19	-.01
CHANGE	.19	.21	.16	.05
COG. STRUC.	.14	.26	-.02	.28
DEFENDENCE	.07	-.14	.10	-.24
DOMINANCE	.48	.63	.26	.37 *
ENDURANCE	.58	.74	.42	.32 *
EXHIBITION	.54	.66	.40	.26
HARMAVOID.	.30	.35	.26	.09
IMPULSIVITY	.34	.36	.31	.05
NURTURANCE	.37	.35	.27	.08
ORDER	.72	.83	.54	.29 **
PLAY	.54	.61	.47	.14
SENTIENCE	.19	.47	.02	.45 *
SOC. RECOG.	.25	.27	.22	.05
SUCCORANCE	.42	.50	.27	.23
UNDERSTAND.	.27	.07	.30	-.23
MEAN	.35	.37	.27	.11

*$p<.05$.

**$p<.01$.

address this latter problem, trait score variances were computed for the subgroups and are presented in Table 6. Of the 15 pairs of trait score correlations having higher values for the consistent groups in Table 5, 10 were based on higher trait score variances for those samples. Of the remaining 5 pairs of correlations having higher values for the variable groups, 2 were based on greater trait score variances for those subgroups. The greatest variance differences, however, were not generally associated with the greatest discrepancies in validity correlations, suggesting that the relationship between imbalances in trait score variances and predictability is not without some source of error. The observed degree of association, nonetheless, raises ambiguities in interpretation which do not exist in studies controlling for the problem (e.g., Bem & Allen, 1974; Zanna et al., 1980). Examination of the trait score frequency distributions revealed that, in general, those trait dimensions having greater dispersion for the behaviourally variable subgroups were skewed (e.g., Achievement, Affiliation, Cognitive Structure, Nurturance, Understanding) with few respondents at one of the two extremes (cf. Kenrick & Braver, 1982, p. 184).

The Kenrick and Stringfield procedure was further applied to the PRF rating data using Bem and Allen's recommended ipsatized variance ratio as the measure of behavioural consistency. Thus, a person's trait consistency score was now computed as that subject's

Table 6

Trait Score Variances for Subjects Above and Below the Median in Self Rated Consistency, for Self Ratings on 20 Adjective Scales

	Consistent subjects (n=45)	Variable subjects (n=45)	
ABASEMENT	5.83	2.58	*
ACHIEVEMENT	1.64	2.78	
AFFILIATION	1.32	2.24	
AGGRESSION	4.89	3.42	
AUTONOMY	2.86	3.07	
CHANGE	6.77	3.77	
COG. STRUC.	1.62	2.65	
DEFENDENCE	3.94	2.92	
DOMINANCE	3.91	3.79	
ENDURANCE	3.85	2.22	
EXHIBITION	5.53	3.54	
HARMAVOID.	6.32	3.70	
IMPULSIVITY	8.76	3.18	**
NURTURANCE	2.20	2.58	
ORDER	5.85	3.58	
PLAY	3.72	2.48	
SENTIENCE	2.81	3.24	
SOC. RECOG.	5.62	3.36	
SUCCORANCE	6.38	2.28	**
UNDERSTAND.	.57	1.61	**
MEAN	4.22	2.95	

*$p<.05$.

**$p<.01$.

response variance across items within the trait scale divided by his/her response variance across all items in the multiscale PRF inventory. The "nomothetic" and "idiographic" correlations are contrasted in the first three columns of Table 7. In comparison to the single item adjective ratings (Table 5), these peer rating validities are generally much higher due to aggregated measurements. The "idiographic" analysis resulted in greater predictability for the consistent versus the variable subgroups for 16 out of the 20 trait dimensions. The same concerns that were expressed with the preceding data based on the adjective ratings (Table 5) are also relevant with these PRF ratings; only four dimensions showed significant differences at or beyond the .05 level, and one trait (Achievement) was actually best predicted by neither consistent nor inconsistent subgroups but by the total sample. Particularly problematic, however, are the rather large discrepancies in trait score variances (Table 8) that result when the ipsatized variance ratio is used to select consistent and inconsistent subsamples. This is to be expected, considering the correlation between the variance ratio measure of consistency and trait extremity (Table 4). High ipsatized variance scores can only occur under the condition of inconsistent item responding, which ultimately produces neither a high nor a low estimate of the respondent's trait level.

Table 7

Differences in Peer Rating Validities for Subjects Above and Below the Median on the Variance Ratio Index of Consistency, for 20 PRF Scales: Consistency and Trait Level Confounded

	All subjects (n=90)	Consistent subjects (n=45)	Variable subjects (n=45)	(difference)
ABASEMENT	.48	.51	.44	.07
ACHIEVEMENT	.63	.60	.62	-.02
AFFILIATION	.57	.63	.48	.15
AGGRESSION	.36	.43	.31	.12
AUTONOMY	.57	.71	.29	.42**
CHANGE	.48	.47	.49	-.02
COG. STRUC.	.51	.57	.38	.19
DEFENDENCE	.34	.43	.16	.27
DOMINANCE	.50	.51	.52	-.01
ENDURANCE	.51	.64	.36	.28
EXHIBITION	.62	.74	.47	.27*
HARMAVOID.	.45	.42	.50	-.08
IMPULSIVITY	.56	.61	.49	.12
NURTURANCE	.61	.62	.56	.06
ORDER	.67	.74	.51	.23
PLAY	.74	.82	.62	.20*
SENTIENCE	.27	.38	.18	.20
SOC. RECOG.	.38	.47	.24	.23
SUCCORANCE	.52	.63	.30	.33*
UNDERSTAND.	.53	.57	.49	.08
MEAN	.52	.58	.42	.15

*$p<.05$.

**$p<.01$.

Table 8

Trait Score Variances for Subjects Above and Below the Median on the Variance Ratio Index of Consistency, for Self Reports on 20 PRF Scales

	Consistent subjects (n=45)	Variable subjects (n=45)	
ABASEMENT	1.07	.35	**
ACHIEVEMENT	1.23	.74	
AFFILIATION	1.60	.89	
AGGRESSION	2.06	.77	**
AUTONOMY	1.36	.43	**
CHANGE	1.03	.54	*
COG. STRUC.	1.37	.52	**
DEFENDENCE	1.37	.67	*
DOMINANCE	2.56	1.01	**
ENDURANCE	1.62	.53	**
EXHIBITION	2.68	1.34	*
HARMAVOID.	3.10	1.06	**
IMPULSIVITY	2.62	.66	**
NURTURANCE	1.42	.74	*
ORDER	3.16	1.39	*
PLAY	1.49	.58	**
SENTIENCE	1.20	.67	
SOC. RECOG.	1.63	.47	**
SUCCORANCE	2.18	.72	**
UNDERSTAND.	1.37	.55	**
MEAN	1.81	.73	

*$p < .05$.

**$p < .01$.

The Bem and Allen procedure. As in the analyses above, subjects were divided into high and low consistency groups for each trait and the peer rating validities contrasted. To control, however, for the inflation and restriction of range problem that emerged using Kenrick's and Stringfield's procedure, subjects were assigned to consistent and variable subgroups at each of the nine discrete categories of the 9-point adjective trait ratings. Thus, following Bem and Allen (1974), all subjects giving a self rating of "9" on the trait of Affiliation were split into two groups based on the median of their ranked consistency scores for that dimension. (Ties at the median were randomly allocated to groups.) The same applied to subjects with a trait rating of "8" and so on to the lowest scorers on the Affiliation scale who had rated themselves as "1" on that dimension. The "consistent" and "variable" subsamples, consequently, had approximately the same trait score means and variances.

The correlations of peer ratings with self ratings on the one-item adjective scales are presented in Table 9, by trait. In column one, the "nomothetic" correlations are shown where all subjects are considered in the computations. Bem and Allen's "idiographic" procedure yields the coefficients of columns 2 and 3 for the consistent and variable subgroups, respectively. Ten of the traits show greater rating validities when considering only subjects who estimate themselves as behaviourally

Table 9

Differences in Peer Rating Validities for Subjects Above and Below the Median on Self Rated Consistency, for 20 Adjective Scales: Consistency and Trait Level Orthogonal

	All subjects (n=90)	Consistent subjects (n=45)	Variable subjects (n=45)	(difference)
ABASEMENT	.17	.00	.31	-.31
ACHIEVEMENT	.28	.33	.24	.09
AFFILIATION	.48	.41	.57	-.16
AGGRESSION	.34	.34	.35	-.01
AUTONOMY	.23	.17	.31	-.14
CHANGE	.19	.26	.13	.13
COG. STRUC.	.14	.35	-.06	.41
DEFENDENCE	.07	.03	.11	-.08
DOMINANCE	.48	.59	.36	.23
ENDURANCE	.58	.61	.56	.05
EXHIBITION	.54	.63	.43	.20
HARMAVOID.	.30	.26	.35	-.09
IMPULSIVITY	.34	.19	.47	-.28
NURTURANCE	.37	.45	.25	.20
ORDER	.72	.72	.72	.00
PLAY	.54	.46	.62	-.16
SENTIENCE	.19	.23	.15	.08
SOC. RECOG.	.25	.12	.40	-.28
SUCCORANCE	.42	.45	.40	.05
UNDERSTAND.	.27	.48	.05	.43 *
MEAN	.35	.35	.34	.02

*$p < .05$.

consistent on the dimensions; the behaviourally variable subjects are not as accurately predicted for these traits. On the other hand, 9 of the traits show the opposite effect--the behaviour of the most variable subjects is predicted most accurately. Only the trait of Understanding demonstrates a significant difference in validity for the two groups ($z = 2.17$, $p<.05$, two-tailed), a difference contrary to the one found previously using Kenrick's and Stringfield's procedure (cf. Table 5). Overall, differences in validity coefficients between the consistent and inconsistent subgroups shrink when the dependency between consistency ratings and trait ratings is accounted for. This is graphically illustrated in Figure 6.

Bem and Allen (1974) used their "ipsatized variance ratio" to divide subjects into consistent/inconsistent subgroups when self-ratings of consistency failed to differentiate predictable from unpredictable subjects for the trait of conscientiousness. This statistic was employed in the present study to identify consistent and inconsistent subjects for the analysis of peer rating validities on the PRF items. The PRF trait scores were first computed as the mean endorsement (9-point ratings) across the 16 items within a scale rounded to the nearest integer. The classification of subjects into consistent and inconsistent subgroupings then proceeded as with the analysis for the single item adjective ratings, occurring at each of the nine trait levels, resulting in groups with

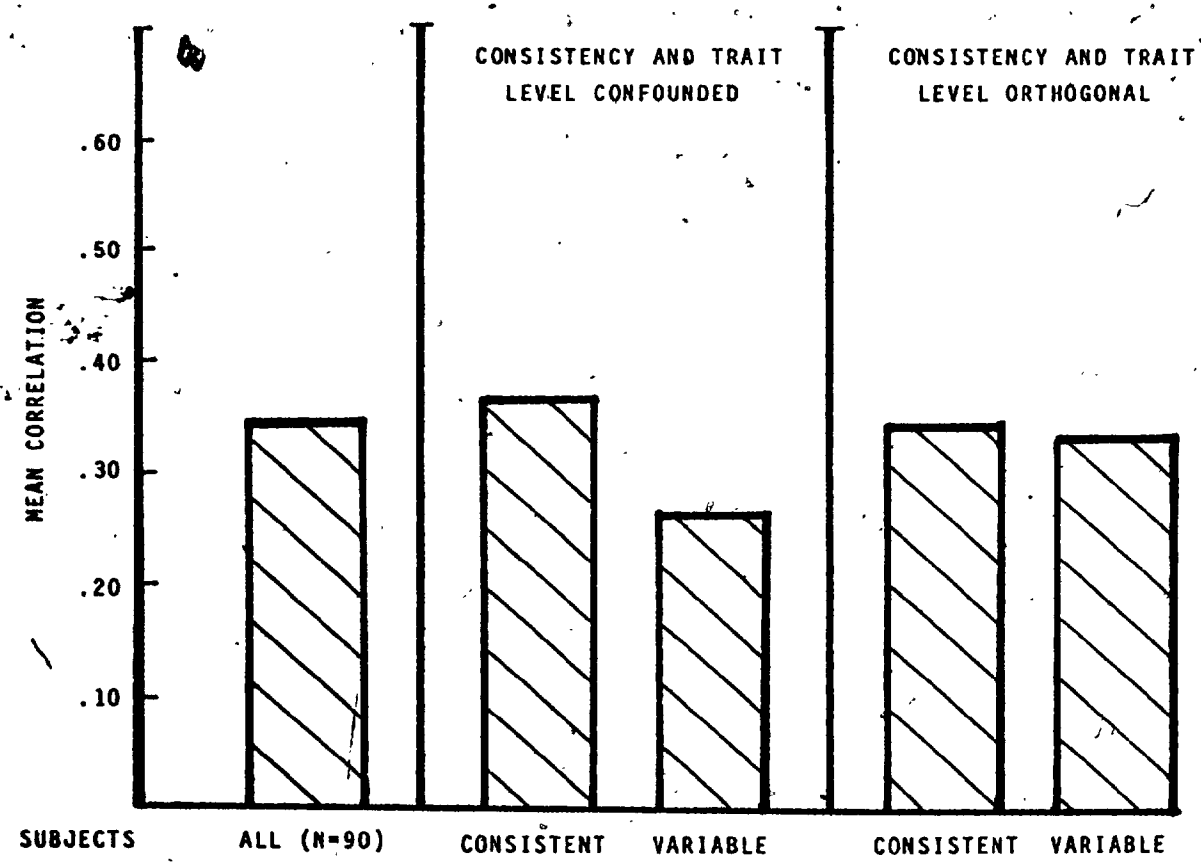

Figure 6. Mean peer rating validities, across 20 adjective scales, for subjects above and below the median on self rated behavioural consistency.

roughly the same trait score means and variances.

The correlations of peer ratings with self ratings on the 16-item PRF scales are presented in Table 10, by trait. Column 1 lists the validity coefficients computed across all subjects, whereas columns 2 and 3 list the correlations for the consistent and inconsistent subgroups, respectively. Segregating subjects into consistent and inconsistent subgroups based on the ipsatized variance ratio yielded, again, asymmetric results. Although 13 of the 20 validities were greater for the consistent groups, 7 were greater for the variable groups. Only two of the differences between correlations were statistically significant, one in the direction opposite to that predicted (Autonomy-- $z = 2.05$, $p<.05$; Harmavoidance-- $z = -2.46$, $p<.05$; two-tailed tests). Again, compared to the coefficients derived using Kenrick and Stringfield's (1980) procedure on the same data (Table 7), a substantial amount of shrinkage is evident in the validity differences when the curvilinearity between consistency measures and trait measures is taken into consideration. This shrinkage is depicted in Figure 7.

Moderated multiple regression analysis. A more general approach to assessing the contribution of a moderator variable in accounting for criterion variance is the moderated multiple regression paradigm discussed in Chapter 2. This linear regression procedure considers the moderator as a continuous variable, such as in the present

Table 10

Differences in Peer Rating Validities for Subjects Above and Below the Median on the Variance Ratio Index of Consistency, for 20 PRF Scales: Consistency and Trait Level Orthogonal

	All subjects (n=90)	Consistent subjects (n=45)	Variable subjects (n=45)	(difference)
ABASEMENT	.48	.46	.50	-.04
ACHIEVEMENT	.63	.61	.64	-.03
AFFILIATION	.57	.63	.48	.15
AGGRESSION	.36	.33	.39	-.06
AUTONOMY	.57	.72	.43	.29 *
CHANGE	.48	.55	.39	.16
COG. STRUC.	.51	.53	.50	.03
DEFENDENCE	.34	.38	.29	.09
DOMINANCE	.50	.41	.59	-.18
ENDURANCE	.51	.55	.47	.08
EXHIBITION	.62	.66	.58	.08
HARMAVOID.	.45	.25	.66	-.41 *
IMPULSIVITY	.56	.61	.51	.10
NURTURANCE	.61	.65	.57	.08
ORDER	.67	.64	.72	-.08
PLAY	.74	.76	.72	.04
SENTIENCE	.27	.30	.22	.08
SOC. RECOG.	.38	.41	.35	.06
SUCCORANCE	.52	.57	.49	.08
UNDERSTAND.	.53	.49	.57	-.08
MEAN	.52	.53	.50	.02

*$p < .05$.

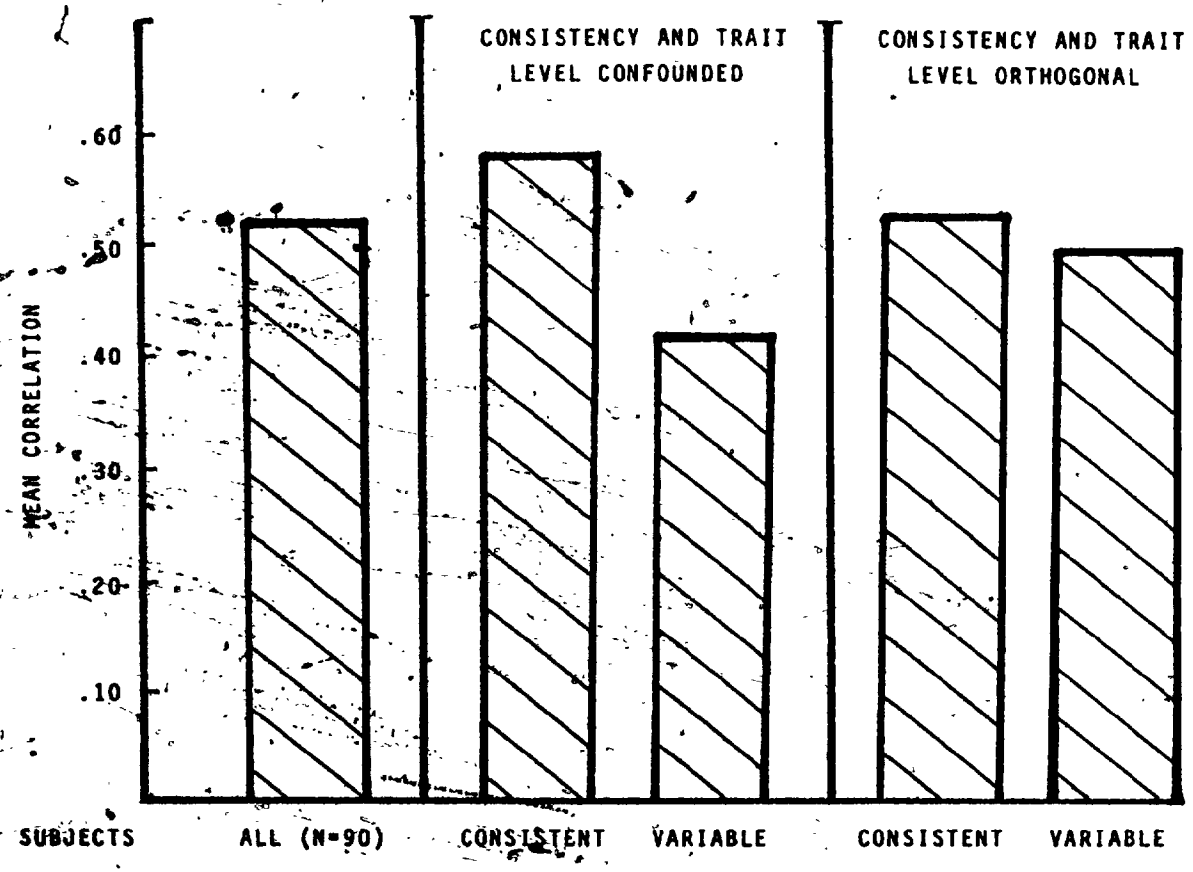

Figure 7. Mean peer rating validities, across 20 PRF scales, for subjects above and below the median on the variance ratio index of consistency.

case, obviating the need to dichotomize a sample into distinct subgroups in assessing differential predictability (Saunders, 1956). Although this procedure was employed by Campus (1974) in an analysis of behavioural consistency as a moderator of predictive accuracy (convergence of sets of self ratings), Bem and Allen (1974) subsequently failed to apply the method to their own data. Recently, however, Tellegen et al. (1982) have emphasized the superiority of moderated regression over the routine median split methodology in searching for moderator variables, suggesting that individual differences in consistency as a moderator of behavioural predictability may show only modest effects, even in a reanalysis of Bem and Allen's own data.

To recapitulate the outline in Chapter 2, the moderated multiple regression model follows a hierarchical analysis (Cohen & Cohen, 1975) where components are sequentially entered into a prediction equation. Criterion Y is regressed on predictor X and the degree of common variance determined. Next, a variable M, thought to moderate the relationship between X and Y, is entered into the equation and evaluated with respect to any increment in the accuracy of predicting Y (i.e., minimizing errors about the response surface). Finally, the moderator effect XM, reflecting the interaction between the predictor and the moderator (Cohen, 1968), is included and evaluated for any improvement in the fit of the response surface to the

observed dependent data. If M moderates the correlation between X and Y, the inclusion of the XM term will result in a nontrivial increase in the multiple correlation of criterion with predictors and will have a regression weight of a sign that is consistent with the expected direction of the moderating effect.

The moderated regression analysis predicting peer adjective ratings with target adjective endorsements using target self rated variability as the moderator is summarized in Table 11, by traits. The rows of F-values reflect the significance of increases in the multiple correlation with the peer rating criterion as the terms are added to the prediction equation, using the test statistic discussed in Chapter 2. It is evident that, in general, the linear component of consistency/variability contributes very little to the prediction of peer ratings after the variance due to the self ratings is partialled from the criterion. Similarly, variability as a moderator does little to minimize the errors of prediction; "bending" the regression plane to accommodate the hypothesized interaction between predictors yields, in most cases, insignificant increments in criterion variance explained. More importantly, the last column of Table 11 shows that the moderator effects, small as they are, are not in a consistent direction. In this analysis, the moderator variable was scored in the direction of variability and the sign of the regression coefficient for the interaction term

Table 11

Summary of Moderated Multiple Regression Analysis of Peer Adjective Ratings on Target Adjective Endorsements (X), Target Self Rated Variability (V), and Their Product (XV), for 20 Trait Dimensions

	F-values[a]				
	F_X	F_V	F_{XV}	Final R	Sign[b]
ABASEMENT	2.51	4.49	0.99	.29	+
ACHIEVEMENT	7.34	15.69	0.06	.47	-
AFFILIATION	26.07	0.38	4.74	.52	-
AGGRESSION	11.76	0.00	0.71	.35	+
AUTONOMY	4.79	2.64	0.06	.28	-
CHANGE	3.40	4.14	0.00	.25	-
COG. STRUC.	1.83	1.16	0.65	.20	+
DEFENDENCE	0.47	3.55	0.03	.21	+
DOMINANCE	26.10	1.82	0.19	.50	-
ENDURANCE	44.28	0.00	0.18	.58	+
EXHIBITION	35.37	1.62	0.68	.55	-
HARMAVOID.	8.94	0.43	0.07	.31	-
IMPULSIVITY	11.11	3.64	0.54	.39	+
NURTURANCE	13.50	0.82	2.13	.40	-
ORDER	93.96	0.63	0.70	.72	-
PLAY	36.39	0.12	1.25	.55	+
SENTIENCE	3.38	0.00	3.66	.28	-
SOC. RECOG.	5.69	0.26	0.24	.26	-
SUCCORANCE	19.30	0.40	1.47	.44	-
UNDERSTAND.	6.99	0.59	2.04	.32	-
MEAN	18.16	2.02	1.02	.39	

Note. $n=90$, $F(1,86)>3.95$, $p<.05$.

[a] F-values, by order of predictor inclusion, reflect the significance in criterion variance accounted for.

[b] Sign of regression coefficient for interaction term.

should be negative if lesser variability is associated with greater degrees of correlation between self and peer ratings of personality (i.e., greater predictability). Six out of 20 of the coefficients, however, are positive in sign and suggest (somewhat) greater predictability for the more variable targets.

The same moderated regression analysis was carried out using mean peer ratings on PRF items, by trait, as the criterion with the targets' mean self endorsements on these items as the initial predictor (Table 12). The moderator variable entered in this regression was the target's ipsatized variance score for each trait. A noticeable difference exists between the first column of F-values of Table 12 and the corresponding column of Table 11, indicating substantially greater predictor/criterion overlap when using aggregated item ratings as variables versus single adjective ratings (cf. the first columns of Table 5 and Table 7). As in the preceding analysis with the adjective ratings, neither the linear variability term nor the variability by trait interaction term yielded consistent and appreciable decreases in residual criterion variance. Furthermore, 11 out of 20 regression coefficients for the interaction term, including the one statistically significant coefficient (Harmavoidance), were positive in sign, suggesting slightly greater degrees of predictability for the more variable subjects on those dimensions.

Table 12

Summary of Moderated Multiple Regression Analysis of Peer PRF Ratings on Target PRF Scores (X), Target Ipsatized Variance Ratio (V), and Their Product (XV), for 20 Trait Dimensions

	F-values[a]				
	F_X	F_V	F_{XV}	Final R	Sign[b]
ABASEMENT	26.38	0.23	0.38	.49	+
ACHIEVEMENT	56.39	0.06	0.00	.63	+
AFFILIATION	41.33	0.29	0.05	.57	-
AGGRESSION	12.80	0.91	0.25	.37	+
AUTONOMY	41.92	3.12	1.16	.60	-
CHANGE	26.05	1.80	0.36	.50	-
COG. STRUC.	31.63	1.56	0.03	.53	-
DEFENDENCE	11.24	0.47	0.06	.34	-
DOMINANCE	29.49	1.03	1.27	.52	+
ENDURANCE	30.18	4.34	0.30	.54	-
EXHIBITION	55.70	0.36	0.02	.62	+
HARMAVOID.	22.03	3.66	5.07	.52	+
IMPULSIVITY	40.18	0.28	0.04	.56	-
NURTURANCE	52.21	0.11	0.01	.61	-
ORDER	71.49	0.18	0.01	.67	-
PLAY	103.78	1.01	3.23	.75	+
SENTIENCE	6.63	0.01	0.03	.27	+
SOC. RECOG.	14.52	2.26	0.01	.40	+
SUCCORANCE	32.58	2.58	0.03	.54	+
UNDERSTAND.	34.54	0.22	1.72	.55	+
MEAN	36.55	1.22	0.70	.53	

Note. $n=90$, $F(1,86)>3.95$, $p<.05$.

[a] F-values, by order of predictor inclusion, reflect the significance in criterion variance accounted for.

[b] Sign of regression coefficient for interaction term.

Tellegen et al. (1982) have argued that regression analyses of the type above may result in the discovery of spurious moderator effects since the calculated consistency index (variance across items) is actually a weighted sum of nonlinear terms (i.e., sums of squares and cross-products of item responses) which must be accounted for separately in the prediction equation. Once these terms are appropriately evaluated using Tellegen's complex hierarchical sequence (p. 97), apparent moderator effects frequently disappear. In the present less conservative analysis only one moderator was found in 20 traits. Another analysis, however, was completed separating the PRF variance index of consistency into its components (see Tellegen et al. 1982, p. 98) and entering them as predictors in the regression sequence. The results are in complete agreement with those of Table 12; only one trait (Harmavoidance) demonstrated a statistically significant moderator effect ($F(1,85) = 7.70$, $p<.01$), which was in the nonhypothesized direction.

Rater Confidence as a Moderator of Judgmental Validity

Overview. One can speculate that the accuracy of subjective estimates of behaviour probabilities is a function of the confidence a judge expresses in making these ratings. Judgments formulated under conditions of uncertainty may represent random guesses or estimates of the base rates of behaviours that show little convergence with other behavioural criteria. Thus, one can consider

rater confidence as a moderator of rating validity, possibly enabling the identification of individual differences in predictive structure.

Using rating confidence as a variable for separating subjects into differentially predictable subgroups may engender a problem equivalent to that discovered with the foregoing application of behavioural consistency. It is possible that confidence in the attributions of traits to targets will increase as a function of the rated extremity of the target on the trait dimension. Extremely high levels of the trait (or its bipolar opposite) would make the dimension a salient characteristic of the target, manifestations of which might be observed by others with less ambiguity than with moderate and nondescript traits. If most confidence is expressed for ratings based on traits that are perceived as being extreme for the target, selecting subgroups of predictable subjects may imply a concomitant inflation of trait score variance for that subgroup and a restricted variance for the remaining (unpredictable) subjects who are rated with low confidence. Differences in validity coefficients between the two groups would be spurious if the median split methodology were employed without consideration of the relationship between rating confidence and trait extremity.

The analyses below determine the dependency of confidence ratings on both the perceived extremity of target traits and the perceived consistency of target

behaviours. Moderated multiple regression analyses were also carried out to determine the extent and direction of any moderating effect of rater confidence on the regression of self on peer judgments of personality.

Confidence and trait extremity. Peer expressions of the confidence of their ratings of target characteristics were correlated with their personality judgments of the target, for each of 20 adjective rating scales. To evaluate the degree of the expected curvilinearity between confidence ratings and personality ratings on bipolar trait dimensions, eta coefficients were computed between these two variables. Finally, trait scores, as estimated by peer judges, were transformed into trait extremity scores by considering a subject's location on a dimension as his/her absolute distance from the point of neutrality on the 9-point adjective rating scales (i.e., "5"). These correlations and eta coefficients are presented in Table 13. Additionally, polynomial regressions were carried out fitting quadratic functions (inverted-U) to the confidence by trait level data. The analyses showed significant ($p<.05$) increments in criterion variance accounted for by the inclusion of the second-order term (square of trait scores) for all 20 dimension.

It is apparent that there is a substantial curvilinear component to the regression of confidence ratings on trait judgments. The more extreme a target is perceived on a dimension of behaviour, the more confident is the judge in

Table 13

Correlations (a) and Eta Coefficients (b) Between Peer Ratings of Confidence and Peer Trait Adjective Ratings, and Correlations (c) of Confidence with Trait Extremity, for 20 Dimensions of Behaviour

	a	b	c
ABASEMENT	-.04	.36	.21
ACHIEVEMENT	.35	.60	.57
AFFILIATION	.36	.71	.57
AGGRESSION	-.24	.65	.47
AUTONOMY	.17	.61	.57
CHANGE	.02	.63	.59
COG. STRUC.	.38	.71	.64
DEFENDENCE	-.11	.56	.49
DOMINANCE	.30	.53	.37
ENDURANCE	.32	.60	.59
EXHIBITION	.23	.61	.44
HARMAVOID.	-.05	.53	.46
IMPULSIVITY	.09	.51	.33
NURTURANCE	.41	.61	.55
ORDER	.08	.62	.61
PLAY	.39	.68	.65
SENTIENCE	.30	.59	.57
SOC. RECOG.	-.04	.53	.47
SUCCORANCE	-.07	.31	.25
UNDERSTAND.	.45	.52	.41

Note. $n=90$, $r>.175$, $p<.05$.

assigning the target to a location on that dimension. As with the measures of behavioural consistency, there was some linear relationship between trait level and confidence ratings for skewed trait score distributions (Table 13, column 1). For example, this sample was generally high in Achievement. This meant few data points were found in the lower left-hand section of the hypothesized inverted-U plot of Achievement by rating uncertainty scores.

It is of interest to note whether or not self ratings of personality are also formulated under different degrees of uncertainty. To address this question, the analyses above were applied to self ratings on the 20 adjective scales and ratings of self confidence in making these judgments. Table 14 mirrors the general patterning of coefficients as found in the previous analyses with peer ratings. Polynomial regressions demonstrated significant ($p<.05$) quadratic components for all but one of the 20 trait dimensions. All second-order terms had regression coefficients in the expected direction. It appears that subjects are not unequivocally certain about their self descriptions but that the uncertainty diminishes as these descriptions reflect their more characteristic dimensions of behaviour.

Confidence and trait consistency. Relations among correlated variables are, in general, not transitive. (Mulaik, 1972, p. 329). Given only the knowledge that variable x is moderately (but imperfectly) correlated with

Table 14

Correlations (a) and Eta Coefficients (b) Between Self Ratings of Confidence and Trait Adjective Endorsements, and Correlations (c) of Confidence with Trait Extremity, for 20 Dimensions of Behaviour

	a	b	c
ABASEMENT	-.02	.55	.37
ACHIEVEMENT	.33	.57	.56
AFFILIATION	.38	.62	.53
AGGRESSION	-.15	.38	.31
AUTONOMY	.42	.49	.44
CHANGE	.24	.45	.38
COG. STRUC.	.58	.66	.63
DEFENDENCE	-.33	.63	.61
DOMINANCE	.37	.64	.61
ENDURANCE	.29	.60	.60
EXHIBITION	.25	.44	.38
HARMAVOID.	.02	.39	.11
IMPULSIVITY	.01	.69	.66
NURTURANCE	.52	.65	.57
ORDER	.32	.68	.62
PLAY	.36	.56	.53
SENTIENCE	.55	.75	.60
SOC. RECOG.	.02	.46	.31
SUCCORANCE	-.10	.46	.43
UNDERSTAND.	.60	.66	.62

Note. $n=90$, $r>.175$, $p<.05$.

variable y and y is moderately correlated with variable z, the precise correlation between x and z is indeterminate, although a range of possible values can be circumscribed. In the present study it was demonstrated that measures of behavioural consistency and of rating confidence are both correlated with trait extremity. It is of interest to determine the extent to which ratings of consistency and confidence are themselves related.

Table 15 column 1 presents the correlations, for 20 adjective scales, of self ratings of consistency on a trait dimension with the confidence expressed in self descriptions using that dimension. The correlation of column 2 represent the association between peer ratings of target consistency and peer confidence in making the trait attributions. These data suggest that target traits that are seen as highly variable in their behavioural manifestations across situations are ascribed to targets (by self and by peers) with relatively low degrees of certainty.

Moderated multiple regression analysis. It has been shown that rater confidence in the attributions of personality characteristics to targets is a function of the perceived salience of those characteristics with respect to individual targets (Table 13 and Table 14). This function contraindicates the use of rating confidence as a moderator of rating validity if the intention is to follow the median split procedure of Kenrick and Stringfield (1980) since, as

Table 15

Correlations of Ratings of Trait Consistency and Judgments of Confidence in Trait Attributions, for 20 Adjective Scales

	Self ratings	Peer ratings
ABASEMENT	.61	.44
ACHIEVEMENT	.44	.61
AFFILIATION	.46	.48
AGGRESSION	.14	.62
AUTONOMY	.23	.51
CHANGE	.48	.44
COG. STRUC.	.30	.48
DEFENDENCE	.45	.51
DOMINANCE	.45	.55
ENDURANCE	.57	.38
EXHIBITION	.33	.42
HARMAVOID.	.26	.31
IMPULSIVITY	.44	.40
NURTURANCE	.62	.50
ORDER	.66	.50
PLAY	.49	.39
SENTIENCE	.56	.42
SOC. RECOG.	.42	.59
SUCCORANCE	.46	.50
UNDERSTAND.	.51	.64
MEAN	.44	.48

Note. $n=90$, $r>.175$, $p<.05$.

with the behavioural consistency moderator variable, the classification of subjects into predictable and unpredictable subgroups based on rater confidence would imply a corresponding classification of subjects into extreme trait and moderate trait subgroups, respectively. The differential trait score variances could yield spurious differential validity coefficients for the two groups. To consider rating confidence as a moderator of rating validity, while controlling for the confounding effects, one could employ Bem and Allen's (1974) median split procedure whereby subjects are divided into high and low confidence groups at several (discrete) levels of the trait. On the other hand, one could pursue a moderated multiple regression analysis where the issue evaporates since the moderator is treated as a continuous rather than dichotomous variable and all subjects are incorporated simultaneously in the analysis. The latter approach was adopted and the results are presented below.

Peer judgments of target personality dimensions were used to predict self ratings of targets on the same dimensions. The peers' expressed confidence in their judgments was used as the variable potentially moderating the relationship between peer and self descriptions of personality. The procedure followed the sequential multiple regression analysis outlined previously. First, self ratings, on each of the 20 personality adjective scales, were regressed on peer ratings of the same

dimensions. The confidence variable was subsequently entered into the equation and evaluated with respect to its linear contribution towards the prediction of the criterion. Finally, the confidence by personality judgment interaction term, representing the moderator effect of interest, was added and examined for nontrivial increases in the multiple correlation of criterion with predictors.

The data in Table 16 suggest that, as with the measures of behavioural consistency, rater confidence does not, in general, have the property of operating as a meaningful moderator of the relationship between different assessments of the same construct. In only two out of 20 cases was there a significant reduction in residual criterion variance as a result of increasing the complexity of the prediction equation through the addition of the interaction term. Although these two effects were in the appropriate direction (i.e., self-peer rating convergences being higher at higher levels of peer rating confidence), 8 out of 20 of the effects were in the opposite direction.

Factor Structure of Ratings of Personality

Overview. The censure of traditional approaches to personality assessment has had several sources. For example, in an early peer rating study Passini and Norman (1966) discovered that the factor structure underlying ratings made by strangers of each other's personality characteristics was highly similar to the structure found when such ratings were formulated by intimate

22

OF / DE

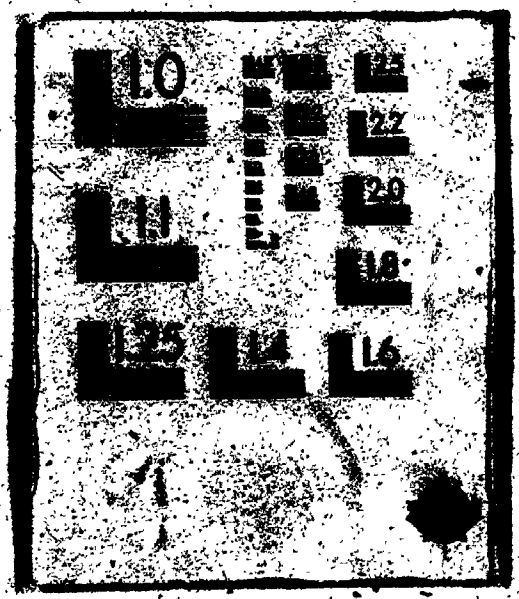

Table 16

Summary of Moderated Multiple Regression Analysis of Self Adjective Endorsements on Peer Adjective Ratings (X), Peer Judgments of Confidence in Adjective Ratings (C), and Their Product (XC), for 20 Trait Dimensions

	F-values[a]				
	F_X	F_C	F_{XC}	Final R	Sign[b]
ABASEMENT	2.51	0.10	5.40	.29	+
ACHIEVEMENT	7.34	0.10	0.03	.28	+
AFFILIATION	26.07	0.26	2.63	.50	+
AGGRESSION	11.76	0.77	0.02	.35	-
AUTONOMY	4.79	0.06	0.09	.23	-
CHANGE	3.40	1.65	0.13	.24	+
COG. STRUC.	1.83	0.43	5.33	.29	+
DEFENDENCE	0.47	0.40	0.41	.12	+
DOMINANCE	26.10	0.10	0.42	.48	+
ENDURANCE	44.28	0.03	0.06	.58	-
EXHIBITION	35.37	0.20	0.00	.54	+
HARMAVOID.	8.94	0.72	0.09	.32	-
IMPULSIVITY	11.11	0.82	0.24	.35	+
NURTURANCE	13.50	0.04	2.34	.40	+
ORDER	93.96	4.98	0.00	.74	+
PLAY	36.39	0.24	1.66	.55	-
SENTIENCE	3.38	5.33	1.39	.33	-
SOC. RECOG.	5.69	0.21	0.03	.25	-
SUCCORANCE	19.30	1.07	2.02	.46	+
UNDERSTAND.	6.99	0.21	0.19	.28	-
MEAN	18.16	0.89	1.12	.38	

Note. $n=90$, $F(1,86)>3.95$, $p<.05$.

[a] F-values, by order of predictor inclusion, reflect the significance in criterion variance accounted for.

[b] Sign of regression coefficient for interaction term.

acquaintances. Previously, Mulaik (1964) and D'Andrade (1965) had noted that even ratings of the similarity in "meaning" of trait descriptive adjectives yielded a factor structure congruent with that emerging from self or peer ratings on the same descriptors. These and other data have prompted the suggestion that peer judgments of personality are not valid since the structure underlying these ratings appears to be invariant under changes in the degree of target-rater acquaintanceship and in the choice of rated object. Raters are thought to rely simply on linguistic conventions or the implied associations among trait descriptors in making behaviour probability judgments rather than predicating such decisions on actual target personalities (D'Andrade, 1965; Shweder, 1975).

With regard to stranger ratings, the discovery of similarities between personality factors derived from peer judgment data and those from self descriptive responses does not in any way imply that peer perceptions are accurate (or inaccurate) in terms of a parallelism between actual and estimated behaviour probabilities. A similarity of structures is a necessary, but not a sufficient, condition for peer rating validity. To illustrate, suppose it is found in this study that the organization of target personality characteristics is highly congruent for peer and self ratings. The peer ratings could then be assigned to targets at random, instead of being assigned to their respective roommates, to simulate stranger ratings. One

would not expect a high degree of association between target item endorsements and peer judgments in this reorganization of the data. The factor structures of the two, however, would not be altered and would remain highly similar to each other.

The following set of analyses was undertaken to determine the factor structures underlying the various ratings of personality. All factor matrices were rotated to a common orientation in the factor space and evaluated for congruence. The "nomothetic" accuracy of judgments of personality was subsequently assessed by the proximity of self determined and peer determined target locations in the factor space.

Before the factor structures were solved, if was decided that the internal consistency reliabilities of the PRF item responses, by targets and peer raters, should be examined. Poor internal consistency of responses might suggest an element of randomness in item endorsements that would advise caution in the interpretation of the underlying structure of those responses. The alpha coefficients of the 16-item PRF scales for the present sample of respondents (self ratings and peer ratings) show levels of internal consistency (Table 17) comparable to those of a normative sample on the slightly longer PRF Form-AA (Jackson, 1974, p. 22). Alpha coefficients were not, of course, available for the one-item adjective scale responses.

Table 17

Alpha Coefficients of Internal Consistency of Self Reports and Peer Ratings for 20 16-Item PRF Scales

	Self (n=90)	Peer (n=90)	PRF-AA[a] (n=202)
ABASEMENT	.69	.79	.63
ACHIEVEMENT	.80	.87	.72
AFFILIATION	.87	.92	.76
AGGRESSION	.83	.85	.76
AUTONOMY	.76	.76	.69
CHANGE	.74	.84	.54
COG. STRUC.	.78	.84	.72
DEFENDENCE	.81	.83	.69
DOMINANCE	.91	.91	.85
ENDURANCE	.83	.87	.78
EXHIBITION	.91	.94	.77
HARMAVOID.	.88	.94	.83
IMPULSIVITY	.88	.91	.67
NURTURANCE	.84	.87	.73
ORDER	.91	.95	.85
PLAY	.79	.90	.69
SENTIENCE	.74	.77	.68
SOC. RECOG.	.81	.84	.80
SUCCORANCE	.87	.88	.78
UNDERSTAND.	.71	.86	.66
MEAN	.82	.87	.73

[a]PRF-AA refers to published internal consistency reliabilities of PRF form-AA with 20 items per scale (Jackson, 1974, p. 22).

Structure of self ratings. Although self ratings were made on both adjective scales and PRF scales, it was decided that the latter inventory would, by virtue of its known discriminant and convergent properties (Jackson, 1974), provide a superior estimate of the personality structure of the respondents. The matrix of PRF scale score intercorrelations was factored by the method of principal components, the eigenvalues of which were used to determine the number of factors to retain for rotation and interpretation. Based on an examination of the "scree line" (Cattell, 1966) and a simultaneous consideration of the "eigenvalue one" criterion (Guttman, 1954), five factors were retained and rotated to simple structure using the orthogonal varimax criterion.

The five orthogonal factors underlying the self ratings of the PRF items are presented in Table 18. These factors account for 69.5% of the total variation among the variables. Factor pattern coefficients having an absolute value greater than or equal to .45 are highlighted and relied on in the interpretation of the structure. The first bipolar dimension might be labelled Impulse Control versus Expression since it is defined by the need for Achievement, Cognitive Structure, Harmavoidance, and Order at one end and Impulsivity and Play at the other. The second factor might be called Independence as Autonomy is contrasted with Nurturance, Succorance, and the need for Social Recognition. Factor three is a dimension related to

Table 18

Rotated Factor Solution for Self Reports of Personality on 20 PRF Dimensions

	I	II	III	IV	V
ABASEMENT	-.02	-.08	-.26	-.79	-.04
ACHIEVEMENT	.54	.35	.17	-.17	.51
AFFILIATION	-.12	-.24	.81	-.18	.04
AGGRESSION	-.32	-.07	.25	.73	-.04
AUTONOMY	-.17	.82	-.07	.18	.15
CHANGE	-.44	.17	.44	.05	.34
COG. STRUC.	.83	-.12	-.11	.04	.05
DEFENDENCE	-.12	.00	-.12	.86	.14
DOMINANCE	.10	.36	.73	.29	.17
ENDURANCE	.37	.42	.18	-.40	.47
EXHIBITION	-.27	-.08	.84	.23	-.06
HARMAVOID.	.67	-.33	-.11	-.10	.04
IMPULSIVITY	-.86	-.07	.04	.24	-.13
NURTURANCE	.14	-.46	.23	-.64	.24
ORDER	.66	-.02	-.06	-.21	-.19
PLAY	-.68	-.12	.53	-.07	-.13
SENTIENCE	-.13	-.42	.05	-.06	.64
SOC. RECOG.	-.01	-.73	-.04	.06	.19
SUCCORANCE	.10	-.82	.06	-.03	-.25
UNDERSTAND.	.06	.15	-.02	.13	.65
% VARIANCE					
TOTAL -	18.4	14.8	13.4	14.2	8.7
COMMON -	26.6	21.3	19.2	20.4	12.5

Social Contact, defined by people high in Affiliation, Exhibition, Play, and Dominance. The traits of Aggression and Defendence are opposed to Abasement and Nurturance on the fourth factor suggesting a continuum related to Aggression versus Passivity. Subjects who scored high on the last factor were those with high scores on the PRF scales of Achievement, Endurance, Understanding, and Sentience. These characteristics appear to describe a domain of behaviours related to Intellectual Orientation.

Self ratings on the 20 one-item adjective scales, designed to measure the PRF constructs, were also intercorrelated and factored by the method of principal components. An examination of the magnitude of the eigenvalues for the respective components suggested five factors for rotation and interpretation. These five dimensions were rotated in the factor space to be maximally coincident with the factors identified for the PRF item responses (Table 18), subject to the constraint of mutual orthogonality (Schonemann, 1966). This type of procedure is known as a linear procrustean transformation, which is a rotation of the factor pattern to the best least-squares approximation to a criterion matrix (Mulaik, 1972, p. 310).

Table 19 presents the five factors of the adjective self rating inventory which account for 58.9% of the total variation among the items. Inspection of the highlighted variables reveals a substantial degree of similarity between the first four factors of the adjective self

Table 19

Rotated Factor Solution for Self Ratings of Personality on 20 Adjective Dimensions

	I	II	III	IV	V
ABASEMENT	.13	-.14	-.23	-.68	-.15
ACHIEVEMENT	.39	.51	.35	.04	.16
AFFILIATION	-.05	-.21	.77	-.22	.11
AGGRESSION	-.25	-.07	.27	.64	.12
AUTONOMY	-.01	.70	-.02	-.04	-.02
CHANGE	-.50	.04	.48	-.03	.12
COG. STRUC.	.62	-.16	.04	-.11	.32
DEFENDENCE	.11	-.21	-.09	.68	-.29
DOMINANCE	.19	.25	.60	.44	.23
ENDURANCE	.50	.49	.22	-.01	.32
EXHIBITION	-.30	-.03	.72	.22	.19
HARMAVOID.	.56	-.45	-.29	.10	.00
IMPULSIVITY	-.65	-.13	.13	.37	.06
NURTURANCE	.16	-.50	.09	-.61	-.01
ORDER	.33	-.07	-.27	-.28	.40
PLAY	-.64	-.02	.49	-.09	-.09
SENTIENCE	-.04	-.44	-.14	-.08	.61
SOC. RECOG.	.26	-.21	.41	-.02	-.46
SUCCORANCE	.06	-.77	-.04	.31	-.11
UNDERSTAND.	.11	.11	.13	-.03	.62
% VARIANCE					
TOTAL -	13.3	12.6	13.2	11.7	8.1
COMMON -	22.6	21.4	22.3	19.9	13.7

ratings and the corresponding dimensions underlying the PRF item responses (Table 18). There seems to be some disparity in the pattern coefficients for the last factor suggesting that the Intellectual Orientation dimension has not been reliably measured. Coefficients of factorial congruence (Harman, 1976, p. 344) were used to assess empirically the similarity between the two structures. These coefficients are shown in column 1 of Table 20 and indicate a high degree of congruence between the first four corresponding dimensions of the factor pattern matrices with substantially less convergence of the two measurements for the last corresponding factor.

Structure of peer ratings. The factorial structure underlying the ratings of targets by roommate peers was determined in order to evaluate any possible disparities between the structure of personality as assessed through self perceptions and the structure that emerges from observer perceptions of target behaviours. If observer ratings of target characteristics are random, or are formulated under conditions of high ambiguity or without relevant personality information, there is no reason to expect a high degree of association between the dimensions latent in these observations and those inferred from self observations.

Mean peer ratings across the items of the PRF scales were intercorrelated and factored by the method of principal components. Five factors were retained and

Table 20

Factor Congruence Coefficients Based on Different Factor Solutions for Ratings of Personality on 20 Traits

	Contrasted ratings of personality		
Factor	Self PRF/ Self adjective	Self PRF/ Peer PRF	Self adjective/ Peer adjective
I -	.94	.96	.87
II -	.92	.97	.92
III -	.91	.97	.90
IV -	.90	.95	.91
V -	.56	.92	.63

rotated, using an orthogonal procrustean transformation, to the criterion matrix based on the factoring of target self reports on the PRF (cf. Table 18). The five factors, accounting for 74.3% of the total variance in the peer ratings, were found to be almost colinear with the factors underlying the self ratings. The coefficients of congruence are presented in column 2 of Table 20 and are equal to or greater than .92. It appears that the perceived organization of personality traits agrees quite closely with the organization based on the decisions of targets concerning their own behaviour probabilities.

Peer ratings on the one-item adjective scales were also factored and five dimensions accounting for 63.0% of the total variance were rotated to the criterion factor pattern matrix of self adjective ratings of Table 19. The factor congruences are shown in column 3 of Table 20 and suggest a high degree of similarity between the first four corresponding factors underlying peers' judgments and self descriptions on the ad hoc one-item rating scales. Only some congruence is evident between the two measurements on the last factor.

Factor structure and judgmental validity. The convergence of self ratings and peer ratings as an index of judgmental validity can be determined by measuring the proximity of actual target and perceived target locations in the factor space defined by the personality variables. This amounts to an evaluation of the degree of association

between each target's factor score and the factor score derived from his/her peer's assessment, for each factor dimension (e.g., Norman & Goldberg, 1966). The factor solutions described in the previous section were used to arrive at factor score estimates (or more appropriately, component scores) for each target on each latent dimension of personality. These estimates were intercorrelated, across targets, for the factors of self ratings and the factors of peer ratings. For example, target scores on the first PRF self rating factor (Table 18) were correlated with their scores on the first PRF peer rating factor. A high degree of peer rating accuracy would be evident in a substantial correlation between self determined and peer estimated factor scores. For the sake of comparison, peer ratings were also assigned randomly to targets in the sample to simulate ratings of target behaviours made by strangers. Such a procedure does not change the factor structure underlying the personality judgments by peers but, in all likelihood, will yield inferior estimates of judgmental accuracy, as assessed by self and peer factor score intercorrelations.

The first column of Table 21 gives the correlations between two different self determined estimates of one's location on the five previously identified factor dimensions. The factor scores used for these correlations were based on the PRF self rating factors (Table 18) and the adjective self rating factors (Table 19). There is

Table 21

Factor Score Intercorrelations Based on Different Factor Solutions for Ratings of Personality on 20 Traits

Factor	Contrasted ratings of personality: Self PRF/ Self adjective	Self PRF/ Peer PRF	Self adjective/ Peer adjective
I -	.80	.70 (.02)[a]	.61 (.07)
II -	.64	.56 (.03)	.46 (.19)
III -	.75	.59 (.12)	.45 (.10)
IV -	.75	.50 (.10)	.48 (.04)
V -	.17	.24 (-.08)	.25 (.04)

Note. $n=90$, $r>.175$, $p<.05$.

[a] Correlations in parentheses based on random pairing of self-peer dyads.

substantial convergence of the two measures for the first four factors with only trivial convergence for the last factor. This analysis, which is analogous to an alternate forms evaluation of reliability, suggests that the Intellectual Orientation factor discussed earlier is not reliably assessed by either the PRF scales, the ad hoc adjective scales, or both. This point was already made in reference to the factor congruences of Table 20.

Column 2 of Table 21 contains the correlations between target self determined factor scores on the PRF and their estimates by roommate peers. Highly significant "nomothetic" validity coefficients are evident for the first four factors with lower but significant rating convergence for the last. Moreover, these values are far in excess of those in parentheses which are related to peer ratings that have been randomly assigned to targets. The same observations generally are true for the factor score convergences of self and peer endorsements of the adjective dimensions of personality (column 3).

It appears that, even for self ratings, the most accurate assessments are on personality variables related to the first factor (Impulse Control versus Expression). Slightly smaller validity coefficients are found for the traits defining the next three factors (Independence, Social Contact, Aggression), and substantially lower values emerge for the last dimension (Intellectual Orientation).

CHAPTER 5

DISCUSSION

Current reviews of psychological research have contributed to a sense of pessimism and frustration regarding the utility of instruments of personality assessment as predictors of behaviour (e.g., Fiske, 1978; Mischel, 1968; Schneider, 1973). The accumulation of failures of personological variables to predict criterion behaviours with statistically acceptable levels of accuracy has not endured without diagnostic analyses. Jackson (1971, 1974, 1982; Jackson & Paunonen, 1980) has proposed several basic requirements for the optimization of personality assessments such as rigorous statistical procedures and stringent criteria for the selection of psychological test items. Disregard of the interaction between persons and situations in the determination of behaviour has been interpreted by Endler and Magnusson (1976; Magnusson & Endler, 1977) as a fundamental defect in many prediction studies. Hogan, DeSoto, and Solano (1977) have recognized other shortcomings of prediction research while Wiggins (1973, 1981) and others (e.g., Kane & Lawler, 1978) have documented numerous examples of the utility of personality predictors of various psychodiagnostic and performance criteria.

A currently popular stream of research in pursuit of situations and operations where relatively reliable and accurate behavioural predictions do occur has been spawned from the idiographic proposition that classes of human behaviour are differentially applicable to individuals (Allport, 1937). Made explicit are the corollaries that only some behavioural domains, or traits, are relevant to or characteristic of any one person, that individuals are relatively invariant across situations in the expression of behaviours related to their characteristic traits, and that the behaviour of these individuals can be predicted with greater precision than the behaviour of persons to whom the trait dimensions do not apply. Recent empirical studies (e.g, Bem & Allen, 1974; Campus, 1974; Cheek, 1982; Kenrick & Braver, 1982; Kenrick & Stringfield, 1980; Turner, 1978; Underwood & Moore, 1981) have purportedly supported this view, demonstrating greater levels of predictor-criterion trait association when measures of behavioural consistency are used to classify subjects as predictable (consistent) and unpredictable (variable).

A peer rating study of personality was designed primarily to address problems identified with some of the idiographic research on consistency as a moderator of the relationship among different assessments of personality that could result in apparent moderator effects not representative of the data. Accounting for these problems permitted an evaluation of the replicability and

generalizability of the widely-recognized findings and conclusions of Bem and Allen (1974). The discussion below examines the implications of the present data with respect to behavioural consistency and the search for individual differences in predictive structure. Secondary findings which bear on some fundamental objections to classical nomothetic assessment approaches are also discussed.

Scaling Behavioural Consistency

Previous research has attempted to define behavioural consistency (or variability) in terms of more than half a dozen different operationalizations while seldom considering questions pertinent to the psychometric validity of the measures. Although some studies simultaneously used more than one of several operationally distinct procedures for identifying individual differences in behavioural consistency, few have attempted to evaluate the convergence of the alternate forms. (A notable exception has been a study by Schneiderman (1980) which found the ipsatized variance and nonipsatized variance indices of consistency to be somewhat related to the observed variability of specific expressive behaviours.) This omission is particularly problematic in Bem and Allen's (1974) analyses since they selectively used a single subjective decision about consistency as a moderator in predicting behaviours related to one dimension (friendliness) while using the ipsatized variance ratio for another dimension (conscientiousness) when the subjective

measure failed to yield significant results. Kenrick and Stringfield (1980) used self, peer, and parent ratings of trait variability but did not report agreement coefficients.

The data presented here show that self ratings of consistency do not agree substantially with estimates by peer roommates (cf. Schutz, 1979). Furthermore, the self rated and ipsatized variance indices used by Bem and Allen are essentially independent, as are the self rated and nonipsatized item response variability measures. The ipsatized variance index of consistency carries the additional problem of confounding within-scale and across-scale item response variability. These observations point to serious ambiguities in the interpretation of these operationalizations as individual difference scales of the same psychological construct and raise questions about the the meaning of published data in the area.

Consistency and Need Strength

It was observed that modern conceptions of behavioural consistency have much in common with traditional characterizations of psychological needs or traits. Intuitively it would seem that individuals high or low on a bipolar dimension of behaviour should, in general, express trait characteristic behaviours with greater regularity than those who assume a middle location on the continuum. Furthermore, it has been demonstrated that there is a mathematical relationship between the maximum variances and

the means of sets of real numbers representing points on a line segment (Figure 1). This latter observation is pertinent to the current issue since such variances and means have been used concurrently to reflect individual differences in trait variability and trait level, respectively (e.g., Bem & Allen, 1974; Underwood & Moore, 1981). The definitional and mathematical analyses above suggest a model in anticipating the cross-situational variability of an individual's trait related behaviours from an awareness of the person's location on the trait dimension: consistency is predicted to increase at the extremes of the trait continuum.

The hypothesized curvilinear relationship between consistency and personality was found for both self ratings and peer ratings. Rushton et al. (1980) have reported similar evidence while Cheek (1982) has not. The ipsatized variance ratio also demonstrated the curvilinear relationship with personality item endorsements. The simple, nonipsatized alternative to the variance ratio also showed some dependency on trait level but less than the other indices of behavioural consistency. Underwood and Moore (1981) have reported no curvilinearity between the variances and means of personality questionnaire responses. As has previously been mentioned, this observation is not unusual in light of the high homogeneity of responses to personality items such as those contained in the PRF (Jackson, 1974) and the consequent restricted range of

variability scores. Alpha coefficients of internal consistency vary inversely as a function of item response variability. The magnitude of these coefficients for the present data reflect individual responses to the items within a trait scale that are relatively uniform. This suggests that the simple variance index of consistency may neither correlate highly with other variables nor reliably discriminate among individuals with respect to the consistency of non-test behaviours.

Consistency as a Moderator

Bem and Allen (1974), following a study by Campus (1974), were able to dichotomize a sample of subjects into groups differing in degree of cross-situational consistency for behaviours related to two independent trait domains. Hypothesizing differential predictability coefficients for the groups, they reported generally greater intercorrelations among different assessments of personality for the behaviourally consistent, in contrast to the more variable, subjects. Others have interpreted this demonstration of the utility of idiographic assumptions of personality as being of great moment and, accordingly, have presented similar demonstrations, with somewhat less than perfect convergence of results (e.g., Cheek, 1982; Kenrick & Braver, 1982; Kenrick & Stringfield, 1980; Lamiell, 1981; Schneiderman, 1980; Underwood & Moore, 1981).

Some of the subsequent studies inspired by Bem and Allen's (1974) work have not strictly followed their original procedure and are, consequently, subject to criticism. Bem and Allen recognized, for example, the possibility of a dependency between trait consistency and personality scores. Because of the dependency, certain experimental procedures may engender statistical confounds: subjects supposedly classified as being "consistent" may in fact be those at the high and low ends of bipolar trait continua while variable subjects may be predominately near the middle. The correlations of personality trait scores with other behavioural criteria could show discrepant values for the consistent and variable subjects simply because of restriction of range effects rather than any real differences in predictive structure. While Bem and Allen precluded the occurrence of the problem in their analyses by ensuring the consistent and variable subgroups had equivalent trait score means and variances, Kenrick and Stringfield (1980) and others (Kenrick & Braver, 1982; Schutz, 1979; Turner, 1978) have not taken suitable precautions. In these studies subjects were divided into groups differing in trait consistency without regard to potential between-group differences in trait level. The re-emergence of the hypothesized differences in the predictive validity of personality assessments for the groups has been interpreted as supporting the idiographic formulations of Bem and Allen.

The data of the present investigation have demonstrated spurious moderator effects when the relationship between personality and consistency is unaccounted for. Orthogonalizing the two variables substantially reduced measured differences in the predictability of the criterion groups, as summarized in Figure 6 and Figure 7. In doing so, the general conclusions of Bem and Allen (1974) were not supported. These findings imply alternative explanations for the data of Kenrick and Stringfield (1980) and Kenrick and Braver (1982) while suggesting that the results of Bem and Allen are not robust and do not generalize beyond the trait dimensions, subject sample, and/or the experimental situation incorporated in their study. Differences between the current findings and those of Bem and Allen cannot be attributed to the exclusive use, in this investigation, of questionnaire measures of personality. Reference has already been made to the predictive utility of the PRF with respect to objectively measured behavioural events (cf. Gifford, 1981, 1982; Meyer & Pepper, 1977) while the relevance of peer ratings to behavioural criteria has been documented in numerous sources (e.g., Jackson, 1974; Kane & Lawler, 1978; Wiggins, 1973).

Although there seems to be a broad consensus on the viability of the hypotheses tested by Bem and Allen, a careful perusal of the present data and other studies reveals a surprising number of specific examples of

failures to replicate or extend their observations. Bem and Allen found self reported consistency to be a useful moderator of the relationship among measures of friendliness. This was also true of the present study when the procedure of moderated multiple regression was followed using the adjective measure of Affiliation. They reported that the ipsatized variance ratio successfully discriminated among predictable and unpredictable subjects on the trait of conscientiousness. Of the personality measures used in this study, perhaps Cognitive Structure and Order best exemplify Bem and Allen's notion of conscientious behaviours. Using the variance ratio index of consistency and following either their median-split procedure or the method of moderated multiple regression failed to provide evidence of significant moderator effects for either dimension.

Other examples of discontinuous results are apparent in the present data and those from other published sources. With respect to the dimension of friendliness, Kenrick and Stringfield (1980) found results to confirm those of Bem and Allen. Underwood and Moore (1981) also reported differential convergences in the predicted direction with measures of sociability but only for peer ratings of target behaviours and not for self ratings. While Cheek (1982) has shown the self reported consistency moderator to be useful for measures of extraversion, this was not the case for a possibly related dimension of agreeableness. Using

moderated multiple regression analyses and a computed variance index of consistency, Campus (1974) failed to find moderator effects for different self report measures of affiliation. Furthermore, Schutz (1979), using a tertile split on self assessed consistency, produced little evidence for a linear relationship between the variability of affiliative behaviours and predictability.

For conscientiousness, Cheek (1982) and Kenrick and Stringfield (1980) presented data supportive of Bem and Allen's results. Schutz (1979) found similar evidence for measures of order but only for peer ratings of targets and not for self ratings. Similarly, Campus (1974) failed to discover moderating effects for self reports on order scales while Schutz also claimed such a failure for the dimension of cognitive structure.

Other disparities can be listed. Campus (1974) used a response variability measure of consistency and moderated regression applied to different self ratings of personality, finding predicted moderator effects for the dimensions of aggression, defendence, deference, exhibition, and sentience. Four of these traits were incorporated in the present study yet none of them demonstrated a significant moderator effect using the same analytic procedure. Turner (1978) showed some supportive evidence for measures of dominance which was not reproduced here nor in the study by Campus. Also of paramount importance is the fact that Tellegen et al. (1982) were

able to apply moderated regression to Bem and Allen's own data, discovering the erosion of apparent moderator effects under the more complex analytic procedure. Finally, curious anomalies have been observed in this study and those of others (Cheek, 1982; Schutz, 1979; Underwood & Moore, 1981) where decreasing degrees of consistency on traits are often associated with increases in predictability.

Critical analyses of the relevant data suggest caution in the unequivocal acceptance of the utility of behavioural consistency as a moderator of personality assessments. More than a few isolated instances of successful hypothesis tests are necessary before the adoption of major modifications to traditional nomothetic assessment strategies (e.g., Lamiell, 1981). Highly replicable empirical observations using sound scientific methodology are primary requisites for a nontrivial assault on modern psychometric theory. Replicable observations have yet to come whereas moderated multiple regression provides a suitable methodology. The median-split procedure followed by Bem and Allen and others has the disadvantages of treating continuous psychological variables as dichotomous, performing hypothesis tests on subsamples of data at the expense of statistical power, and not making salient the significance of increments in predictability due to moderating effects over the values derived using conventional nomothetic computations. These deficiencies

can be overcome by the use of moderated multiple regression analysis, which also has the advantage of conceptual clarity. Campus (1974) was the first to employ this technique proposed by Saunders (1956) to investigate the consistency issue. Unfortunately others, including Bem and Allen (1974), have failed to follow her lead.

With respect to the search for replicable results and the use of moderated multiple regression, it should be noted that there is, as observed by Tellegen et al. (1982), a large accumulation of convincing research testifying to the lack of robustness of complex terms in regression equations. Wiggins (1973) presents empirical evidence supporting his argument that increases in the complexity and number of terms in multinomial expressions may initially yield increases in predictive accuracy, but these predictive gains are often more than lost under cross-validation of the regression coefficients. Simple linear regression equations have shown the least shrinkage in multiple correlations of predictor batteries with their respective criteria. Significant regression weights for interaction terms in moderated multiple regression applications may tend not to replicate on new samples of data.

Confidence as a Moderator.

It was inferred that the degree of confidence expressed by a rater in formulating behaviour probability judgments might be a moderator of the accuracy of those

judgments. Subjective decisions about which raters have little confidence may reflect unverified suppositions that are not predictive of other behavioural criteria. Rater uncertainty, however, was observed to be related to the perceived extremity of target traits and, consequently, shares problems with the variable of behavioural consistency. To classify personality ratings as predictive (made with high confidence) and not predictive (low confidence) would be, simultaneously, to classify the ratings as extreme and moderate, respectively. Differential personality trait score variances for the predictable and unpredictable groups could contribute illegitimately to moderator effects.

Since the relationship between predictor and moderator variables causes no restriction of range problems for moderated multiple regression, this procedure was applied to the personality adjective ratings. Specifically examined was the moderating effect of personality rating confidence by peers on the convergence of peer-target personality judgments. Only two of 20 trait dimensions showed the hypothesized effect (Abasement and Cognitive Structure). For the others, the correlations between peer and self ratings were essentially constant at all levels of rater uncertainty. Although a person may express low confidence in a personality attribution, it appears that such a judgment is as accurate, or inaccurate, as one that is less speculative.

Aggregated Measurements

It has often been argued that the pervasiveness of insubstantial coefficients of validity for many measures of personality compels the adoption of idiographic assessment strategies (e.g., Bem & Allen, 1974; Lamiell, 1981; Mischel, 1968). Much of the concern over the inadequacy of personality predictors of behaviour, however, is due to poorly formulated research designs that nurture inordinately high levels of measurement error. For example, many empirical studies have attempted to scale individual differences on personality dimensions by using single, subjective decisions recorded on bipolar multipoint rating scales. Such one-item measures of psychological attributes are typically highly unreliable, being replete with error variance. Since the reliability of a measure determines the upper limit of its validity (Nunnally, 1978, p. 220), any procedure which reduces error variance would be desirable. As frequently observed, psychometric theory provides a simple solution through a consideration of the reliability of composites of scores or aggregated measurements (Epstein, 1979, 1980; Harris, 1980; Jackson & Paunonen, 1980; Kenrick & Braver, 1982; Rushton et al., 1981). Increments in the number of replicated measurements reduce the error of measurement, increase reliability, and raise the level of the validity ceiling.

In responding to criticisms of one-item global measures of personality attributes, Kenrick and Braver (1982) have argued that such individual decisions "can reflect careful implicit pooling of (self or other) observations" (p. 183). Under these conditions the judgments could be highly reliable and valid in predicting behaviour. They report a mean split-half reliability of .55 averaged over 16 two-item scales with a meah retest reliability of .62.

For the present data, the 20 single-item adjective scales constituted global trait decisions and their endorsements conceivably reflected the latent pooling of numerous target behaviours. Although the reliability of the adjective ratings was not established, their validity was. The average correlation between peer and self ratings on these items was .35 (Table 5) over 20 dimensions. Using the 16-item PRF aggregated personality scores yielded an average alpha reliability of .82 for self ratings and .87 for peer ratings (Table 17). These reliabilities, which correspond to the average of all possible split-half coefficients (Cronbach, 1951), are substantially higher than those reported by Kenrick and Braver (1982) for their much shorter scales where "implicit averaging" was inferred. More importantly, the PRF aggregated measurements yielded greater validities in comparison to the bipolar adjective scales for all but two trait dimensions (Endurance and Order), with a mean self-peer

correlation of .52 (Table 7). Implicitly averaged ratings on global adjective dimensions appear to result in reliabilities and validities that are subordinate to those derived from the explicit averaging of individual behaviour-situation probability judgments. Moreover, the data of the present investigation lend suppprt to the construct validity of the PRF.

Personality Structure

The claim has been made that assessments of personality (nomothetic or idiographic) may lack external validity due to semantic conventions which bias judgments of behaviour. This issue has its origin in the findings of a number of studies of the factorial structure underlying trait ratings. Referring to a variety of sources, Wiggins (1973, ch. 8) has illustrated an impressive stability to the measured interrelatedness of trait dimensions. Ratings collected on essentially the same 20 bipolar adjective scales by several investigators have yielded a highly invariant factor structure across diverse samples of respondents and judgmental situations. For example, Norman (1963) repeatedly found the same five orthogonal factors using several sets of peer ratings of close acquaintances on the adjective scales. This would not be unusual except for the fact that even peer ratings of complete strangers produced essentially the same factors (Passini & Norman, 1966). Moreover, D'Andrade (1965) instructed raters to judge the "similarity of meaning" of all pairs of the 20

dimensions and found an underlying structure highly congruent with that of Norman's peer ratings.

This observed invariance in the perceived interrelatedness of trait descriptors has been construed as damaging to the notion that peer ratings can demonstrate empirical validity since the correlational structure is not affected by target-rater familiarity nor by choice of rated object. Citing D'Andrade (1965), Mischel (1968) has maintained that "so-called traits at least in part exist as components of the verbal terms used to describe the external world; they do not necessarily mirror the external world itself" (p. 46). Schneider (1973) has commented that judgments of personality could arguably be "heavily influenced by linguistic factors" (p. 301).

It must be assumed that these indictments refer to both nomothetic and idiographic assessment formulations where semantically-based subjective decisions about personality constitute the fundamental data. But the fact that the perceived organization of personality is the same for both acquaintance and stranger ratings does not imply that these trait inferences have the same validity in the sense of a correspondence between perceived and actual target behaviour probabilities. When self ratings of personality are available, factor analysis can be used to establish the validity of the trait attributions of others by evaluating the association between self determined and other determined locations on the factor dimensions (e.g.,

Norman & Goldberg, 1966).

The observation in this study of similar factor structures underlying both self reported and peer estimated probabilities of behaviour was seen as a prerequisite for the validity of the peers' judgments. (To have found different structures for the two sets of ratings would have been to preclude appreciable levels of rating validity since self and peer emplacements of targets would be along dissimilar factor dimensions.) Factor scores provided estimates of respondent locations along the dimensions of the factor space. These estimates by self and by roommate judges showed substantial convergence in the present data indicating some evidence for validity in the peer attributions of traits to targets. Furthermore, when simulating stranger ratings of targets by randomly pairing the roommate dyads, although the factor structures of self and other personality ratings did not change with this reordering of the data, the validities of the ratings failed to differ significantly from zero.

As far as the factor structure of ratings of word meaning is concerned, one would expect an approximation to the structure of person ratings if the linguistic connections and implied associations among trait descriptors are "a consequence of inductive generalizations of real-world events" (Jackson & Paunonen, 1980, p. 534). It is difficult to conceive of universally learned relationships among personality characteristics that are

only illusory, as has been suggested elsewhere (e.g., Shweder, 1975).

Conclusions

A peer rating study of personality was designed to evaluate empirically several problems identified with current research and theorizing involving the pervasive belief that the predictive validity of behavioural assessments is moderated by individual differences in cross-situational behaviour consistency. Other issues pertinent to the psychometric validity of instruments of personality measurement were also examined.

Based on a critical analysis of relevant literature and an interpretive integration of the outcomes of this investigation with data from other sources, the following conclusions are warranted:

1. Customary procedures for scaling individual differences in behavioural consistency based on distinct operations show poor levels of convergent validity. Even the same operations do not yield acceptable convergence, as when self ratings are compared to peer ratings.

2. Individual trait consistency is related to the measured level of the trait such that greater consistency is observed at the extremes of bipolar dimensions of behaviour. This occurs whether consistency is assessed by indices of response variability or by the subjective intuitions of raters.

3. Failing to consider the relationship between measures of consistency and personality in the identification of subjects as predictable in their behaviours (consistent) or unpredictable (variable) can result in the discovery of moderator effects that are spurious. Inflated and attenuated trait score variances for the consistent and variable subsamples, respectively, increase the likelihood of between-group differences in the correlation of personality measures with other behavioural criteria.

4. Accounting for the relationship between consistency and personality scores by experimental or statistical procedures can reduce observed differences in the predictability of differentially consistent individuals to such an extent that the utility of behavioural consistency as a moderator of the validity of personality assessments is questioned. Additionally, numerous examples of inconsistencies found in several published sources testify to the dubious value of a reorganization of traditional assessment strategies in pursuit of an idiographic ideology.

5. Moderated multiple regression is conceptually and statistically the most meaningful and effective treatment of data when evaluating psychological variables as moderators. Among its advantages is an immunity to false moderator effects that may surface when a sample is divided into consistent and variable halves without regard to between-group differences in trait score distributions.

6. The degree of confidence expressed by a rater in formulating a personality judgment is a function of the perceived salience of the characteristic being rated with respect to the target. This is true of both peer attributions and self attributions of trait dimensions. Rater confidence, furthermore, appears not to moderate the accuracy of trait perceptions.

7. The arithmetic averaging of independent judgments of behaviour probabilities results in estimates of personality that are psychometrically superior to implicitly averaged behaviour probabilities reflected in global trait judgments. Aggregated measurements yield levels of reliability and criterion validity considerably greater than those of single observations.

8. The organization of personality traits reflected in the factor structure underlying ratings of behaviour demonstrates substantial stability. This stability is maintained across different raters of personality (self, other) and different methods of assessment.

9. Peer judgments of target behaviour probabilities are highly reliable and show statistically significant measures of validity when target self reports are used as a criterion. This occurs at both the individual trait and the trait composite or factorial levels and may be a function of target-rater familiarity.

REFERENCES

Abelson, R. P. Sex differences in predictability of college grades. Educational and Psychological Measurement, 1952, 12, 638-644.

Allport, G. W. Personality: A psychological interpretation. New York: Holt, 1937.

Bem, D. J., & Allen, A. On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. Psychological Review, 1974, 81, 506-520.

Bem, D. J., & Funder, D. C. Predicting more of the people more of the time: Assessing the personality of situations. Psychological Review, 1978, 85, 485-501.

Campus, N. Transituational consistency as a dimension of personality. Journal of Personality and Social Psychology, 1974, 29, 593-600.

Cattell, R. B. The scree test for the number of factors. Multivariate Behavioral Research, 1966, 1, 245-276.

Cheek, J. Aggregation, moderator variables, and the validity of personality test: A peer-rating study. Journal of Personality and Social Psychology, 1982, in press.

Cohen, J. Multiple regression as a general data-analytic system. Psychological Bulletin, 1968, 70, 426-433.

Cohen, J., & Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. New York: Erlbaum, 1975.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

D'Andrade, R. G. Trait psychology and componential analysis. American Anthropologist, 1965, 67, 215-228.

Endler, N. S., & Magnusson, D. Toward an interactional psychology of personality. Psychological Bulletin, 1976, 83, 956-974.

Epstein, S. The stability of behavior: I. On predicting most of the people much of the time. Journal of Personality and Social Psychology, 1979, 37, 1097-1126.

Epstein, S. The stability of behavior: II. Implications for psychological research. American Psychologist, 1980, 35, 790-806.

Fiske, D. W. Strategies for personality research. San Francisco: Jossey-Bass, 1978.

Frederiksen, N., & Melville, S. D. Differential predictability in the use of test scores. Educational and Psychological Measurement, 1954, 14, 647-656.

Gifford, R. Sociability: Traits, settings, and interactions. Journal of Personality and Social Psychology, 1981, 41, 340-347.

Gifford, R. Affiliativeness: A trait measure in relation to single-act and multiple-act behavioral criteria. Journal of Research in Personality, 1982, 16, 128-134.

Glass, G. V., & Stanley, J. C. Statistical methods in education and psychology, Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Green, P. E., & Carroll, J. D. Mathematical tools for applied multivariate analysis, New York: Academic Press, 1976.

Guttman, L. Some necessary conditions for common factor analysis. Psychometrika, 1954, 19, 149-161.

Harman, H. H. Modern factor analysis. Chicago: University of Chicago Press, 1976.

Harris, J. G. Nomovalidation and idiovalidation: An expedition to chart and verify the personality profile. American Psychologist, 1980, 35, 729-744.

Hogan, R., DeSoto, C. B., & Solano, C. Traits, tests, and personality research. American Psychologist, 1977, 32, 255-264.

Jackson, D. N., The dynamics of structured personality tests: 1971. Psychological Review, 1971, 78, 229-248.

Jackson, D. N. Personality Research Form manual. Port Huron, Mich.: Research Psychologists Press, 1974.

Jackson D. N. Some preconditions for valid person perception. In M. P. Zanna, E. T. Higgins, & C. P. Herman (Eds.), Consistency in social behavior: The Ontario symposium, Vol. 2. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1982.

Jackson, D. N., & Guthrie, G. M. A multitrait-multimethod evaluation of the Personality Research Form. Proceedings of the 76th Annual Convention of the American Psychological Association, 1968, 177-178.

Jackson, D. N., & Paunonen, S. V. Personality structure and assessment. Annual Review of Psychology, 1980, 31, 503-551.

Kane, J. S. & Lawler, E. E. Methods of peer assessment. Psychological Bulletin, 1978, 85, 555-586.

Kellogg, R. L. The Strong Vocational Interest Blank as a differential predictor of engineering grades. Educational and Psychological Measurement, 1968, 28, 1213-1217.

Kenrick, D. T., & Braver, S. L. Personality: Idiographic and nomothetic! A rejoinder. Psychological Review, 1982, 89, 182-186.

Kenrick, D. T., & Stringfield, D. O. Personality traits and the eye of the beholder: Crossing some traditional philosophical boundaries in the search for consistency in all of the people. Psychological Review, 1980, 87, 88-104.

Lamiell, J. T. Toward an idiothetic psychology of personality. American Psychologist, 1981, 36, 276-289.

Magnusson, D., & Endler, N. S. Interactional psychology: Present status and future prospects. In D. Magnusson & N. S. Endler (Eds.), Personality at the crossroads. Hillsdale, N. J.: Erlbaum, 1977.

Meyer, J. P., & Pepper, S. Need compatibility and marital adjustment in young married couples. Journal of Personality and Social Psychology, 1977, 35, 331-342.

Mischel, W. Personality and assessment. New York: Wiley, 1968.

Mulaik, S. A. Are personality factors raters' conceptual factors? Journal of Consulting Psychology, 1964, 28, 506-511.

Mulaik, S. A. The foundations of factor analysis. New York: McGraw-Hill, 1972.

Nisbett, R. E., Caputo, C., Legant, P., & Maracek, J. Behavior as seen by the actor and as seen by the observer. Journal of Personality and Social Psychology, 1973, 27, 154-164.

Norman, W. T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. Journal of Abnormal and Social Psychology, 1963, 66, 574-583.

Norman, W. T., & Goldberg, L. R. Raters, ratees, and randomness in personality structure. Journal of Personality and Social Psychology, 1966, 4, 681-691.

Nunnally, J. C. Psychometric theory. New York: McGraw-Hill, 1978.

Passini, F. T., & Norman, W. T. A universal conception of personality structure? Journal of Personality and Social Psychology, 1966, 4, 44-49.

Pervin, L. A. Personality: Theory, assessment and research. New York: Wiley, 1970.

Rothstein, M., & Jackson, D. N. Decision making in the employment interview: An experimental approach. Journal of Applied Psychology, 1980, 65, 271-283.

Rushton, J. P., Jackson, D. N., & Paunonen, S. V. Personality: Nomothetic or idiographic? A response to Kenrick and Stringfield. Psychological Review, 1981, 88, 582-589.

Saunders, D. R. Moderator variables in prediction. Educational and Psychological Measurement, 1956, 16, 209-222.

Scheier, M. F., Buss, A. H., & Buss, D. M. Self-consciousness, self-report of aggressiveness, and aggression. Journal of Research in Personality, 1978, 12, 133-140.

Schneider, D. J. Implicit personality theory: A review. Psychological Bulletin, 1973, 79, 294-309.

Schneiderman, W. A personality dimension of consistency versus variability without the use of self-reports or ratings. Journal of Personality and Social Psychology, 1980, 39, 158-164.

Schonemann, P. H. A generalized solution of the orthogonal procrustes problem. Psychometrika, 1966, 31, 1-16.

Schutz, L. E. A multi-method assessment of idiographic properties of the Personality Research Form. Unpublished Master of Science thesis, University of Kentucky, 1979.

Shweder, R. A., How relevant is an individual difference theory of personality? Journal of Personality, 1975, 43, 455-485.

Sorrentino, R. M., & Short, J. C. The case of the mysterious moderates: Why motives sometimes fail to predict behavior. Journal of Personality and Social Psychology, 1977, 35, 478-484.

Stones, M. J., & Burt, G. Quasi-statistical inference in rating behavior: I. Preliminary investigation. Journal of Research in Personality, 1978, 12. 381-389.

Stricker, L. J. Compulsivity as a moderator variable: A replication and extension. Journal of Applied Psychology, 1966, 50, 331-335.

Tellegen, A., Kamp, J., & Watson, D. Recognizing individual differences in predictive structure. Psychological Review, 1982, 89, 95-105.

Turner, R. G. Consistency, self-consciousness, and the predictive validity of typical and maximal personality measures. Journal of Research in Personality, 1978, 12, 117-132.

Turner, R. G., & Gilliam, B. J. Identifying the situationally variable subject: Correspondence among different self-report formats. Applied Psychological Measurement, 1979, 3, 361-369.

Turner, R. G., & Gilliland, L. The comparative relevance and predictive validity of subject generated trait descriptions. Journal of Personality, 1979, 47, 230-244.

Underwood, B., & Moore, B. S. Sources of behavioral consistency. Journal of Personality and Social Psychology, 1981, 40, 780-785.

Wiggins, J. S. Personality and prediction: Principles of personality assessment. Reading. Mass.: Addison-Wesley, 1973.

Wiggins, J. S. Clinical and statistical prediction: Where are we and where do we go from here? Clinical Psychology Review, 1981, 1, 3-18.

Zanna, M. P., Olson, J. M., & Fazio, R. H. Attitude-behavior consistency: An individual difference perspective. Journal of Personality and Social Psychology, 1980, 38, 432-440.

APPENDIX A

INSTRUCTIONS FOR PERSONALITY RESEARCH FORM AND ACQUAINTANCE RATINGS

Peer Rating Instructions

Your name and student number - _____

Full name of person being rated - _____

Your sex - _____

Sex of person being rated - _____

In the attached "Statement Booklet" you will find a series of statements which a person might use to describe him or herself. Read each statement and decide how characteristic the statement is of the person you have been asked to rate (target person). Use the attached 9-point rating scales to record your responses.

Consider the example below:

"I enjoy all kinds of vigorous hobbies"

9 - Extremely characteristic of the target
(8) - Very characteristic of the target
7 - Moderately characteristic of the target
6 - Slightly characteristic of the target
5 - Neutral
4 - Slightly uncharacteristic of the target
3 - Moderately uncharacteristic of the target
2 - Very uncharacteristic of the target
1 - Extremely uncharacteristic of the target

The response indicates that the statement is very characteristic of the person being rated. Your own answer might have been different.

Please respond to all of the items in the "Statement Booklet" and record your answers on the attached rating forms. In marking your answers, be sure that the number of the statement you have just read is the same as the number on the answer sheet.

Please do not omit any item even if you are not completely sure of your answer. Before you begin, fill in the information requested at the top of this page and the Degree of Acquaintance Rating below.

How well are you acquainted with the person you are rating?
(circle the appropriate number)

7 - We are the best of acquaintances and spend a great deal of time together.

6 - The person and I are very well acquainted and we interact socially quite often.

5 - I know the person quite well and we interact socially on occasion.

4 - I interact with the person at work or class but not socially.

3 - The person is a passing acquaintance whom I know by reputation and the odd casual conversation.

2 - I know the person by reputation only and I do not know him/her well enough to talk to.

1 - I do not know the person at all. I may have heard the name but I know nothing about him/her.

APPENDIX B

MATERIALS FOR ADJECTIVE RATINGS OF TRAITS, VARIABILITY, AND CONFIDENCE

PERSONALITY SELF RATINGS

Your name and student number _____

On the following pages, you will find a series of trait adjectives that could be used to describe a person. After each adjective is a definition of the term. These definitions identify behaviours that are characteristic of individuals representing the traits. Also given after each adjective is the opposite trait.

Your task is to read each trait label and corresponding definition carefully, then make the following 3 judgments with respect to your own behaviour. Example:

ANXIOUS (opposite to RELAXED)

Tends to worry over inconsequential matters; is tense, restless, uneasy; apprehensive about the future; is easily upset.

A. How characteristic is ANXIOUS or RELAXED of you?

ANXIOUS is Extremely Characteristic	9	8	7	6	Neutral 5	4	3	2	(1)	RELAXED is Extremely Characteristic

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely Variable	9	8	7	6	Average 5	4	3	2	(1)	Extremely Consistent

C. How confident are you of judgment A?

Extremely Confident	9	(8)	7	6	Average 5	4	3	2	1	Not at all Confident

---

Notice that the rater has estimated ANXIOUS to be extremely uncharacteristic of him/her and the opposite trait (RELAXED) to be extremely characteristic. The second rating indicates that he/she is extremely consistent in behaviours related to being relaxed and calm across various situations. The third rating represents very high confidence in making the first judgment on how characteristic the trait is.

Please continue on to the following pages reading the descriptions carefully and recording your judgments about yourself by circling the appropriate numbers on the 9-point rating scales.

---

1. HUMBLE (opposite to ARROGANT)

Shows a high degree of humility; accepts blame and criticism even when not deserved; exposes him/herself to situations where he/she is in an inferior position; tends to be self-effacing.

A. How characteristic is HUMBLE or ARROGANT of you?

HUMBLE is Extremely Characteristic	9	8	7	6	5	4	3	2	1	ARROGANT is Extremely Characteristic
					Neutral					

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely Variable	9	8	7	6	5	4	3	2	1	Extremely Consistent
					Average					

C. How confident are you of judgment A?

Extremely Confident	9	8	7	6	5	4	3	2	1	Not at all Confident
					Average					

---

2. ASPIRING (opposite to UNAMBITIOUS)

Aspires to accomplish difficult tasks; maintains high standards and is willing to work toward distant goals; enjoys competition; always strives to attain excellence.

A. How characteristic is ASPIRING or UNAMBITIOUS of you?

ASPIRING is Extremely Characteristic	9	8	7	6	5	4	3	2	1	UNAMBITIOUS is Extremely Characteristic
					Neutral					

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely Variable	9	8	7	6	5	4	3	2	1	Extremely Consistent
					Average					

C. How confident are you of judgment A?

Extremely Confident	9	8	7	6	5	4	3	2	1	Not at all Confident
					Average					

---

3. SOCIABLE (opposite to WITHDRAWN)

Enjoys being with friends and people in general; accepts people readily; makes efforts to win friendships and maintain personal ties.

A. How characteristic is SOCIABLE or WITHDRAWN of you?

SOCIABLE is Extremely Characteristic	9	8	7	6	5	4	3	2	1	WITHDRAWN is Extremely Characteristic
					Neutral					

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely Variable	9	8	7	6	5	4	3	2	1	Extremely Consistent
					Average					

C. How confident are you of judgment A?

Extremely Confident	9	8	7	6	5	4	3	2	1	Not at all Confident
					Average					

---

4. AGGRESSIVE (opposite to EVEN-TEMPERED)

Enjoys combat and argument; easily annoyed; sometimes willing to hurt people to get his/her way; may seek to "get even" with people who are perceived as having harmed him/her.

A. How characteristic is AGGRESSIVE or EVEN-TEMPERED of you?

AGGRESSIVE is Extremely Characteristic	9	8	7	6	5	4	3	2	1	EVEN-TEMPERED is Extremely Characteristic
					Neutral					

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely Variable	9	8	7	6	5	4	3	2	1	Extremely Consistent
					Average					

C. How confident are you of judgment A?

Extremely Confident	9	8	7	6	5	4	3	2	1	Not at all Confident
					Average					

---

5. INDEPENDENT (opposite to DEPENDENT)

Tries to break away from restraints, confinement, or restrictions of any kind; enjoys being unattached, free, not tied to people, places, or obligations.

A. How characteristic is INDEPENDENT or DEPENDENT of you?

INDEPENDENT is Extremely Characteristic	9	8	7	6	5	4	3	2	1	DEPENDENT is Extremely Characteristic
					Neutral					

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely Variable	9	8	7	6	5	4	3	2	1	Extremely Consistent
					Average					

C. How confident are you of judgment A?

Extremely Confident	9	8	7	6	5	4	3	2	1	Not at all Confident
					Average					

---

6. CHANGEABLE (opposite to PREDICTABLE)

Likes new and different experiences; dislikes routine and avoids it; may readily change opinions or values in different circumstances; adapts readily to changes in environment.

A. How characteristic is CHANGEABLE or PREDICTABLE of you?

CHANGEABLE is Extremely Characteristic	9	8	7	6	5	4	3	2	1	PREDICTABLE is Extremely Characteristic
					Neutral					

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely Variable	9	8	7	6	5	4	3	2	1	Extremely Consistent
					Average					

C. How confident are you of judgment A?

Extremely Confident	9	8	7	6	5	4	3	2	1	Not at all Confident
					Average					

---

7. EXACTING (opposite to IMPRECISE)

Does not like ambiguity or uncertainty in information; wants all questions answered completely; desires to make decisions based on definite knowledge, not on guesses.

A. How characteristic is EXACTING or IMPRECISE of you?

EXACTING is										IMPRECISE is
Extremely					Neutral					Extremely
Characteristic	9	8	7	6	5	4	3	2	1	Characteristic

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely										Extremely
Variable	9	8	7	6	5	4	3	2	1	Consistent

C. How confident are you of judgment A?

Extremely										Not at all
Confident	9	8	7	6	5	4	3	2	1	Confident

---

8. DEFENSIVE (opposite to OPEN)

Readily suspects that people are against him/her; ready to defend him/herself at all times; takes offense easily; does not accept criticism readily.

A. How characteristic is DEFENSIVE or OPEN of you?

DEFENSIVE is										OPEN is
Extremely					Neutral					Extremely
Characteristic	9	8	7	6	5	4	3	2	1	Characteristic

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely										Extremely
Variable	9	8	7	6	5	4	3	2	1	Consistent

C. How confident are you of judgment A?

Extremely										Not at all
Confident	9	8	7	6	5	4	3	2	1	Confident

---

9. ASSERTIVE (opposite to SUBMISSIVE)

Attempts to control his/her environment, and to influence or direct other people; expresses opinions forcefully; enjoys the role of leader and may assume it spontaneously.

A. How characteristic is ASSERTIVE or SUBMISSIVE of you?

ASSERTIVE is Extremely Characteristic	9	8	7	6	5	4	3	2	1	SUBMISSIVE is Extremely Characteristic
					Neutral					

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely Variable	9	8	7	6	5	4	3	2	1	Extremely Consistent
					Average					

C. How confident are you of judgment A?

Extremely Confident	9	8	7	6	5	4	3	2	1	Not at all Confident
					Average					

---

10. PERSISTENT (opposite to GIVES UP EASILY)

Willing to work long hours; doesn't give up quickly on a problem; persevering, even in the face of great difficulty; patient and unrelenting in his/her work habits.

A. How characteristic is PERSISTENT or GIVES UP EASILY of you?

PERSISTENT is Extremely Characteristic	9	8	7	6	5	4	3	2	1	GIVES UP EASILY is Extremely Characteristic
					Neutral					

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely Variable	9	8	7	6	5	4	3	2	1	Extremely Consistent
					Average					

C. How confident are you of judgment A?

Extremely Confident	9	8	7	6	5	4	3	2	1	Not at all Confident
					Average					

---

11. EXHIBITIONISTIC (opposite to SHY)

Wants to be the center of attention; enjoys having an audience; engages in behaviour which wins the notice of others; may enjoy being dramatic or witty.

A. How characteristic is EXHIBITIONISTIC or SHY of you?

EXHIBITIONISTIC is										SHY is
Extremely					Neutral					Extremely
Characteristic	9	8	7	6	5	4	3	2	1	Characteristic

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely					Average					Extremely
Variable	9	8	7	6	5	4	3	2	1	Consistent

C. How confident are you of judgment A?

Extremely					Average					Not at all
Confident	9	8	7	6	5	4	3	2	1	Confident

---

12. CAUTIOUS (opposite to ADVENTURESOME)

Does not enjoy exciting activities, especially if danger is involved; avoids risk of bodily harm; seeks to maximize personal safety.

A. How characteristic is CAUTIOUS or ADVENTURESOME of you?

CAUTIOUS is										ADVENTURESOME is
Extremely					Neutral					Extremely
Characteristic	9	8	7	6	5	4	3	2	1	Characteristic

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely					Average					Extremely
Variable	9	8	7	6	5	4	3	2	1	Consistent

C. How confident are you of judgment A?

Extremely					Average					Not at all
Confident	9	8	7	6	5	4	3	2	1	Confident

---

13. IMPULSIVE (opposite to PATIENT)

Tends to act on the "spur of the moment" and without deliberation; gives vent readily to feelings and wishes; speaks freely; may be volatile in emotional expression.

A. How characteristic is IMPULSIVE or PATIENT of you?

IMPULSIVE is										PATIENT is
Extremely					Neutral					Extremely
Characteristic	9	8	7	6	5	4	3	2	1	Characteristic

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely					Average					Extremely
Variable	9	8	7	6	5	4	3	2	1	Consistent

C. How confident are you of judgment A?

Extremely					Average					Not at all
Confident	9	8	7	6	5	4	3	2	1	Confident

---

14. SYMPATHETIC (opposite to INSENSITIVE)

Gives sympathy and comfort; assists others whenever possible; interested in caring for children, the disabled, or the infirm; readily performs favours for others.

A. How characteristic is SYMPATHETIC or INSENSITIVE of you?

SYMPATHETIC is										INSENSITIVE is
Extremely					Neutral					Extremely
Characteristic	9	8	7	6	5	4	3	2	1	Characteristic

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely					Average					Extremely
Variable	9	8	7	6	5	4	3	2	1	Consistent

C. How confident are you of judgment A?

Extremely					Average					Not at all
Confident	9	8	7	6	5	4	3	2	1	Confident

---

15. ORGANIZED (opposite to DISORDERLY)

Concerned with keeping personal effects and surroundings neat and organized; dislikes clutter, confusion, lack of organization.

A. How characteristic is ORGANIZED or DISORDERLY of you?

ORGANIZED is Extremely Characteristic	9	8	7	6	5	4	3	2	1	DISORDERLY is Extremely Characteristic
					Neutral					

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely Variable	9	8	7	6	5	4	3	2	1	Extremely Consistent
					Average					

C. How confident are you of judgment A?

Extremely Confident	9	8	7	6	5	4	3	2	1	Not at all Confident
					Average					

---

16. FUN-LOVING (opposite to SERIOUS)

Does many things "just for fun"; spends a good deal of time participating in games, sports, social activities, and other amusements; enjoys jokes, funny stories; is easy-going.

A. How characteristic is FUN-LOVING or SERIOUS of you?

FUN-LOVING is Extremely Characteristic	9	8	7	6	5	4	3	2	1	SERIOUS is Extremely Characteristic
					Neutral					

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely Variable	9	8	7	6	5	4	3	2	1	Extremely Consistent
					Average					

C. How confident are you of judgment A?

Extremely Confident	9	8	7	6	5	4	3	2	1	Not at all Confident
					Average					

---

17. SENTIENT (opposite to UNPERCEPTIVE)

Notices smells, sounds, sights, tastes, and the way things feel; remembers these sensations and believes that they are an important part of life; is sensitive to many forms of experience.

A. How characteristic is SENTIENT or UNPERCEPTIVE of you?

SENTIENT is										UNPERCEPTIVE is
Extremely					Neutral					Extremely
Characteristic	9	8	7	6	5	4	3	2	1	Characteristic

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely					Average					Extremely
Variable	9	8	7	6	5	4	3	2	1	Consistent

C. How confident are you of judgment A?

Extremely					Average					Not at all
Confident	9	8	7	6	5	4	3	2	1	Confident

---

18. STATUS-SEEKING (opposite to UNPRETENTIOUS)

Desires to be held in high esteem by acquaintances; concerned about reputation and what other people think of him/her; works for the approval and recognition of others.

A. How characteristic is STATUS-SEEKING or UNPRETENTIOUS of you?

STATUS-SEEKING is										UNPRETENTIOUS is
Extremely					Neutral					Extremely
Characteristic	9	8	7	6	5	4	3	2	1	Characteristic

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely					Average					Extremely
Variable	9	8	7	6	5	4	3	2	1	Consistent

C. How confident are you of judgment A?

Extremely					Average					Not at all
Confident	9	8	7	6	5	4	3	2	1	Confident

---

19. SEEKS SYMPATHY (opposite to SELF-SUFFICIENT)

Frequently seeks the sympathy, protection, love, advice, and reassurance of other people; may feel insecure or helpless without such support; confides difficulties readily to others.

A. How characteristic is SEEKS SYMPATHY or SELF-SUFFICIENT of you?

SEEKS SYMPATHY is										SELF-SUFFICIENT is
Extremely					Neutral					Extremely
Characteristic	9	8	7	6	5	4	3	2	1	Characteristic

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely					Average					Extremely
Variable	9	8	7	6	5	4	3	2	1	Consistent

C. How confident are you of judgment A?

Extremely					Average					Not at all
Confident	9	8	7	6	5	4	3	2	1	Confident

---

20. CURIOUS, ANALYTICAL (opposite to UNREFLECTIVE)

Wants to understand many areas of knowledge; values synthesis of ideas, logical thought, particularly when directed at satisfying intellectual curiosity.

A. How characteristic is CURIOUS, ANALYTICAL or UNREFLECTIVE of you?

'CURIOUS,' ANALYTICAL is										UNREFLECTIVE is
Extremely					Neutral					Extremely
Characteristic	9	8	7	6	5	4	3	2	1	Characteristic

B. To what extent do behaviours related to the trait dimension above vary from situation to situation for you?

Extremely					Average					Extremely
Variable	9	8	7	6	5	4	3	2	1	Consistent

C. How confident are you of judgment A?

Extremely					Average					Not at all
Confident	9	8	7	6	5	4	3	2	1	Confident

APPENDIX C

PRF SCALE MEANS FOR SELF RATINGS

PRF SCALE MEANS FOR PEER RATINGS

ADJECTIVE SCALE MEANS FOR SELF RATINGS OF TRAITS, VARIABILITY, AND CONFIDENCE

ADJECTIVE SCALE MEANS FOR PEER RATINGS OF TRAITS, VARIABILITY, AND CONFIDENCE

MEAN 9-POINT SELF RATINGS FOR 16 PRF-E ITEMS, BY TRAIT

	Males (n=20)	Females (n=70)	Total (n=90)
AB	4.59	4.69	4.67
AC	5.40	5.77	5.69
AF	6.01	6.38	6.30
AG	5.07	4.67	4.76
AU	5.03	4.66	4.74
CH	5.70	5.86	5.82
CS	5.51	5.50	5.50
DE	5.03	4.54	4.65
DO	5.98	5.53	5.63
EN	5.04	5.45	5.36
EX	5.75	5.68	5.70
HA	4.58	5.24	5.10
IM	4.93	4.40	4.52
NU	5.27	6.38	6.14
OR	5.00	5.23	5.18
PL	5.74	5.66	5.68
SE	5.32	6.03	5.87
SR	5.28	5.69	5.60
SU	4.50	5.08	4.95
UN	5.06	5.14	5.12

MEAN 9-POINT PEER RATINGS FOR 16 PRF-E ITEMS, BY TRAIT

	Males (n=20)	Females (n=70)	Total (n=90)
AB	4.42	4.50	4.48
AC	5.24	5.43	5.39
AF	6.07	6.19	6.16
AG	4.89	4.54	4.62
AU	5.03	4.63	4.72
CH	5.19	5.16	5.16
CS	5.29	5.54	5.48
DE	5.06	5.01	5.02
DO	5.37	5.15	5.20
EN	5.38	5.15	5.20
EX	5.78	5.59	5.63
HA	4.61	5.52	5.32
IM	4.72	4.62	4.64
NU	4.76	6.03	5.75
OR	4.89	5.06	5.03
PL	5.73	5.58	5.62
SE	5.07	5.83	5.66
SR	4.84	5.50	5.36
SU	4.49	5.19	5.04
UN	4.83	4.97	4.94

MEAN 9-POINT ADJECTIVE SELF RATINGS, BY TRAIT (N=90).

	Trait Ratings	Variability Ratings	Confidence Ratings
AB	5.14	3.51	7.31
AC	7.23	3.22	7.88
AF	7.61	3.39	7.97
AG	3.64	3.60	7.66
AU	6.54	4.22	7.50
CH	6.18	3.61	7.49
CS	6.87	3.38	7.53
DE	3.68	3.68	7.39
DO	6.36	3.81	7.63
EN	6.72	3.90	7.54
EX	6.06	4.40	7.60
HA	4.44	4.36	7.48
IM	5.23	3.87	7.44
NU	7.34	3.03	7.63
OR	6.50	3.13	7.99
PL	7.20	3.40	7.83
SE	6.86	3.63	7.14
SR	6.23	3.70	7.53
SU	4.73	4.32	7.20
UN	7.42	3.53	7.50

MEAN 9-POINT ADJECTIVE PEER RATINGS, BY TRAIT (N=90)

	Trait Ratings	Variability Ratings	Confidence Ratings
AB	5.00	3.24	7.22
AC	6.77	3.34	7.39
AF	7.34	3.19	7.93
AG	3.93	3.21	7.64
AU	5.86	3.59	7.27
CH	5.19	3.77	6.84
CS	6.50	3.34	7.04
DE	4.52	3.29	7.24
DO	5.79	3.36	7.36
EN	6.53	3.68	7.26
EX	6.06	3.36	7.64
HA	4.73	3.62	7.29
IM	4.89	3.54	7.29
NU	6.83	3.02	7.46
OR	6.00	3.13	7.84
PL	6.94	3.14	7.91
SE	6.38	3.64	6.36
SR	5.63	3.46	6.89
SU	4.53	3.14	7.17
UN	6.36	3.33	7.17

APPENDIX D

SUMMARIES OF POLYNOMIAL REGRESSIONS OF TARGET VARIABILITIES ON TRAIT LEVEL

SUMMARY OF POLYNOMIAL REGRESSIONS OF TARGET VARIABILITY ON TRAIT LEVEL, FOR 20 DIMENSIONS:

SELF RATED VARIABILITY BY SELF ADJECTIVE RATINGS (X)

	b_X	b_X^2	b_0	R
AB	1.75 **	-0.18 **	-0.07	0.38
AC	1.40 *	-0.16 **	2.03	0.56
AF	1.17	-0.15 *	3.30	0.53
AG	2.29 **	-0.22 **	-0.81	0.57
AU	1.16	-0.12	2.23	0.29
CH	1.74 **	-0.16 **	0.03	0.38
CS	-0.18	-0.02	5.73 *	0.35
DE	2.29 **	-0.21 **	-1.11	0.58
DO	1.20	-0.13 *	2.05	0.38
EN	1.72 **	-0.19 **	1.68	0.54
EX	1.75 *	-0.17 **	0.73	0.33
HA	1.85 **	-0.19 **	0.78	0.41
IM	2.32 **	-0.22 **	-1.04	0.53
NU	1.36 *	-0.17 **	2.56	0.66
OR	1.59 **	-0.18 **	1.23	0.59
PL	1.82 *	-0.17 **	-0.19	0.38
SE	0.88	-0.12 *	3.68 *	0.53
SR	1.59 **	-0.14 **	-0.09	0.29
SU	2.71 **	-0.26 **	-1.42	0.52
UN	1.13	-0.14	3.23	0.57

PEER RATED VARIABILITY BY PEER ADJECTIVE RATINGS (X)

AB	0.92 *	-0.10 *	1.53	0.27
AC	1.49 *	-0.16 **	1.22	0.48
AF	2.29 **	-0.21 **	-1.21	0.46
AG	2.11 **	-0.21 **	-0.85	0.56
AU	2.24 **	-0.23 **	-0.54	0.60
CH	1.85 **	-0.17 **	-0.37	0.41
CS	1.27 **	-0.14 **	1.56	0.49
DE	2.63 **	-0.26 **	-1.97 **	0.68
DO	1.16 **	-0.13 **	1.53	0.40
EN	1.66 **	-0.16 **	0.48	0.38
EX	1.69 **	-0.17 **	0.11	0.46
HA	1.62 **	-0.17 **	0.64	0.42
IM	1.54 **	-0.14 **	0.67	0.41
NU	1.98 **	-0.20 **	-0.30	0.55
OR	1.72 **	-0.17 **	-0.16	0.43
PL	1.67 **	-0.16 **	-0.12	0.38
SE	1.75 **	-0.19 **	0.68	0.48
SR	1.68 **	-0.17 **	0.20	0.39
SU	1.63 **	-0.16 **	-0.18	0.39
UN	0.68	-0.09 *	3.22 **	0.46

Note. *$p<.05$, **$p<.01$; for all R, $p<.05$.

SUMMARY OF POLYNOMIAL REGRESSIONS OF TARGET VARIABILITY ON TRAIT LEVEL, FOR 20 DIMENSIONS:

IPSATIZED VARIANCE BY PRF SELF REPORTS (X)

	b_X	b_{X^2}	b_0	R
AB	0.48 **	-.033 **	-0.86 *	0.41 **
AC	0.25 **	-.016 **	-0.20	0.44 **
AF	0.32 **	-.019 **	-0.63	0.54 **
AG	0.36 **	-.023 **	-0.54	0.44 **
AU	0.44 **	-.028 **	-0.87 *	0.46 **
CH	0.47 **	-.028 **	-1.25 *	0.44 **
CS	0.43 **	-.028 **	-0.77	0.51 **
DE	0.38 **	-.022 **	-0.93 **	0.47 **
DO	0.35 **	-.022 **	-0.70 *	0.61 **
EN	0.43 **	-.028 **	-0.90 **	0.57 **
EX	0.29 **	-.020 **	-0.25	0.61 **
HA	0.52 **	-.031 **	-1.12 **	0.58 **
IM	0.35 **	-.022 **	-0.65 **	0.55 **
NU	0.37 **	-.023 **	-0.55	0.62 **
OR	0.29 **	-.018 **	-0.36	0.51 **
PL	0.41 **	-.025 **	-0.84	0.48 **
SE	0.35 *	-.023 *	-0.24	0.42 **
SR	0.59 **	-.035 **	-1.55 **	0.64 **
SU	0.34 **	-.021 **	-0.62 *	0.48 **
UN	0.48 **	-.029 **	-0.93	0.38 **

PRF SCALE VARIANCES BY PRF SELF REPORTS (X)

AB	1.01	-.066	0.80	0.12
AC	0.39	-.029	2.55	0.14
AF	1.11	-.065 *	-0.98	0.27 *
AG	0.95	-.059	0.42	0.14
AU	0.35	-.033	3.25	0.17
CH	1.14	-.069	-0.74	0.16
CS	0.67	-.056	2.86	0.29 *
DE	0.52	-.024	0.91	0.14
DO	1.18 *	-.077 **	-1.05	0.38 **
EN	1.73 *	-.112 **	-2.60	0.31 *
EX	1.05 *	-.075 **	0.54	0.43 **
HA	1.91 **	-.109 **	-3.41	0.37 **
IM	0.93 *	-.058	-0.02	0.22
NU	-0.03	-.016	5.60	0.33 **
OR	0.91 *	-.058 *	0.31	0.23
PL	0.51	-.035	2.43	0.14
SE	0.53	-.048	4.40	0.26 *
SR	2.12 **	-.129 **	-4.01	0.31 *
SU	0.46	-.022	1.39	0.16
UN	2.29 *	-.139 **	-3.76	0.28 *

Note. *$p<.05$, **$p<.01$.

END

18 01 83

FIN